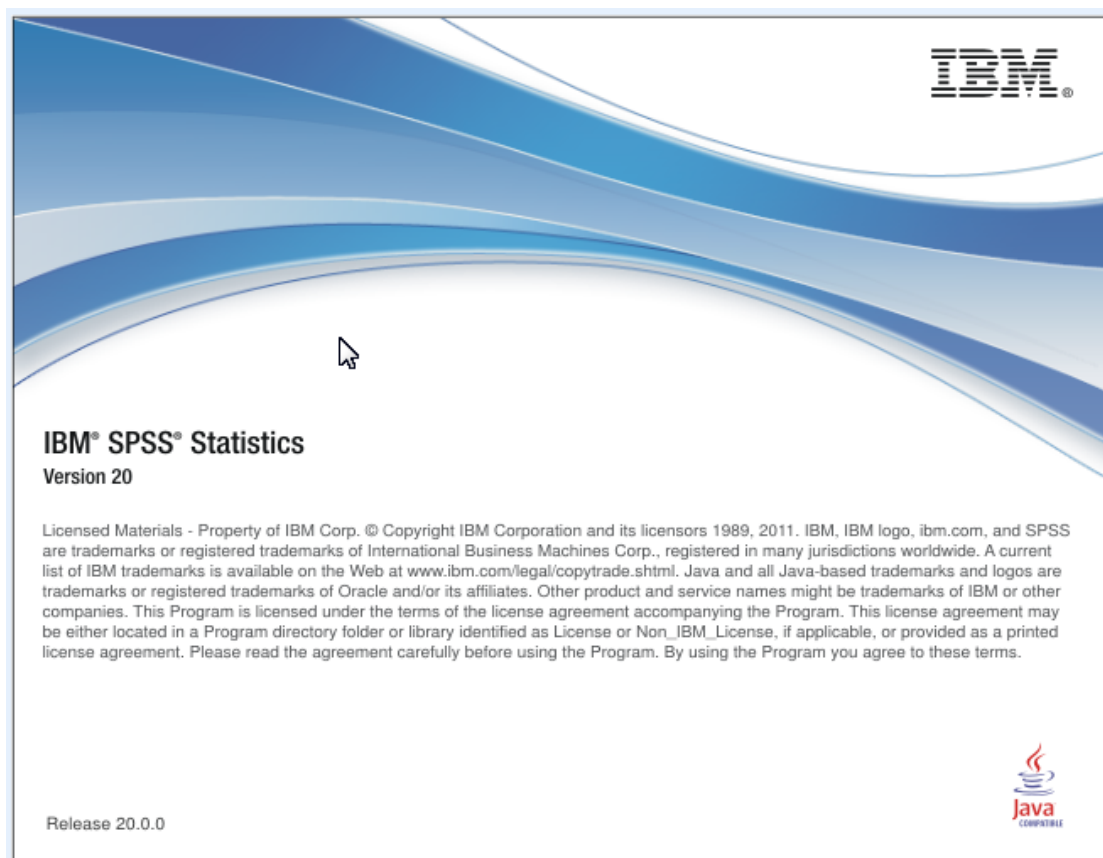


Notes on using IBM SPSS



www.som.soton.ac.uk/learn/resmethods/software/SPSSforWindows/

Contents

1 Introduction.....	1
Conventions used in these notes.....	1
Medicine Coding Guide	3
2 Running SPSS for Windows.....	5
SPSS windows.....	5
SPSS menus.....	6
SPSS toolbars	7
The Status Bar	7
The SPSS Viewer	8
Using the SPSS Help System	9
3 Defining variables using the Data Editor.....	10
4 SPSS rules for defining variables	10
Variables Names.....	10
Type of variable.....	11
Width.....	11
Decimals.....	11
Label and Values	12
Missing.....	13
Columns.....	13
Align.....	13
Measure	14
Role	14
5 Setting up variables in SPSS.....	14
6 Saving the Data Editor Information.....	18
Copying an existing variable to create a new variable.....	18
Copying Variable Attributes	20
7 Entering Data	21
Changing the width of a column	21
Problem you might encounter.....	22
Key combinations for moving around the data editor.....	22
Displaying Values or Value Labels.....	22
Correcting data	23
Replacing values.....	23
Modifying values	23
Other Data Editor Operations.....	23
Cut, Copy and Paste Data Values	23
Add and Delete Cases	23
Add and Delete Variables.....	24
Copying Variable settings	24
Change the order of (moving) variables	24
Sorting the Variable view.....	24
Multiple Data Views	25
8 Checking Data.....	27
9 Viewing Output.....	27
Checking Variable Definitions.....	28
Further Checking	28
10 Printing Output.....	30

11 Saving Output	30
12 Summarising Data.....	31
Categorical (nominal, ordinal) data.....	32
Continuous (scale, interval, ratio) data.....	32
Crosstabulations	36
13 Why all this pasting? - the SPSS Syntax file	36
Saving and Printing the Syntax window	36
Retrieving and Editing the Syntax file	37
14 Creating New Variables and Modifying Data	38
15 The Calculator Pad.....	39
Symbols and their meanings.....	39
Functions	39
Missing Values and Functions	40
The Date and Time Wizard	40
Counting	42
Tips and Hints.....	43
IF - Combining Logical and Relational operators.....	43
Recode.....	44
16 Listing Data Values.....	46
17 Saving modified data	46
18 Charts	47
19 Bar charts	47
Clustered bar chart.....	49
20 Scatterplots - XY graphs	50
21 Incorporating SPSS charts into Word or PowerPoint.....	51
22 Saving Charts	51
23 Modifying Charts	52
Chart Editor Menu options	53
Modifying Chart Elements	54
Modifying a scatter plot	57
Adding a regression line.....	57
24 Comparing Means and Medians	58
Two Independent Samples: t-test	59
Comparing Two Medians: Mann-Whitney test.....	60
Paired samples t-test.....	61
Non parametric Paired Samples: Wilcoxon	62
25 Comparing Proportions 2x2 tables.....	63
Chi-squared test for a 2x2 table.....	64
Interpreting the output	64
Yates' continuity correction	65
Fisher's exact test	65
Other statistics from Crosstabs.....	65
Paired Data (McNemar's test)	65
The kappa (κ) statistic	65
26 Help with Statistical Procedures	66
27 Further Reading	66

28 Analysing Two or more Groups - Split Files.....	67
29 Case selection – Filtering.....	68
30 Sorting Cases	70

1 Introduction

These notes are designed to take you through the stages of analysing a set of data using SPSS for Windows (referred to as SPSS in this document). The notes are written to give you an idea of how SPSS can analyse and present data. They are not a replacement for the SPSS manuals. Only a selection of features will be described here. These will include:

- Describing the variables used to SPSS
- Entering data
- Simple analysis and checking of the data
- Modification of the data and computation of new variables
- Statistical analysis of the data
- Production of graphical output

We have based these notes on a questionnaire; however SPSS can be used with any set of data. The questionnaire is equivalent to a data collection form, or pages from a laboratory notebook. The example used is based on data taken from a larger study on 'Health and Lifestyles'. Variables have been selected to demonstrate how to enter and analyse different types of data using SPSS. The questions selected were re-ordered to present the information here in a coherent fashion

Conventions used in these notes

Several different typefaces have been used in these notes to distinguish actions and information. The conventions followed are:

File Open	This represents a sequence of commands from a windows menu. Select the File menu and then click on the Open option from the drop down menu.
User Input	This is text you will enter.
Variable Name	The name of a text box, or panel in a window
Type	A Windows command button.
Click	Click (or Double click) the left mouse button.
<Delete>	Press the key marked Delete. (Do not press < or >).
<Ctrl A>	With the Ctrl key pressed down, press the A key.
Note	A useful tip or feature.
Syntax	SPSS program code that appears in a syntax window.

All Answers are Confidential

PLEASE TICK THE APPROPRIATE BOX OR BOXES AND FILL IN THE SPACES WHERE INDICATED

- Date of Interview _____/_____/_____
1. What is your date of Birth? _____/_____/_____
2. What sex are you?
 Male 1 Female 2 Missing 9
3. What is your marital status?
 Single (never married) 1
 Married or living as married 2
 Widowed 3
 Divorced 4
 Separated 5
 Missing 9
4. What is your height (without shoes)? _____ feet _____ inches
5. What is your weight (lightly dressed)? _____ kilograms
6. Do you have any long-standing illness, disability or infirmity that has trouble you over a period of time, or that is likely to affect you over a period of time?
 Yes 1 No 2 Not sure 3 Missing 9
- If **Yes**, are you registered disabled?
 Yes 1 No 2 Not applicable 8 Missing 9
7. Which of the following reasons prevent you from taking more exercise?
 (Please tick **no more** than **three** boxes)
- Lack of leisure time 01
 - Lack of money 02
 - Lack of transport 03
 - Lack of easily available facilities at work 04
 - Lack of interesting or relevant activities 05
 - Illness or disability 06
 - Lack of incentive 07
 - Don't believe in exercise 08
 - Other reasons 09

Office Use Only

ID

DOI

DOB

GENDER

MSTATUS

FEET INCHES

•
WEIGHT

LSILL

REGDISAB

NOEX1

NOEX2

NOEX3

Note: The questionnaire assumes that the answers will be coded after completion. Names below each set of boxes in the 'Office Use Only' column identify the variables that you will create. It is a good idea to create a coding guide or scheme from the data collection form (shown on the next page). This assists in the description of the data to SPSS. As this is a self-coding form, boxes for GENDER, LSILL, REGDISAB are not included in the Office Use Only column.

Medicine Coding Guide

File _____ Sheet _____ of _____
 Project Title: _____ Date _____
 Name of Investigator _____ Record _____





Variable Name	Type (Measurement)	Variable label (Recommended max of 40 characters)	Range	Value Labels (Recommended max of 20 characters)	Missing and N/A
<i>ID</i>	<i>Numeric 4 (Scale)</i>	<i>Case ID</i>	<i>1 - 5000</i>		
<i>DOI</i>	<i>Date dd-mmm-yyyy (Scale)</i>	<i>Date of Interview</i>	<i>1-Jun-2004 - 31-Oct-2004</i>		<i>1-Jan-1900</i>
<i>DOB</i>	<i>Date dd-mmm-yyyy (Scale)</i>	<i>Date of Birth</i>	<i>1-Jan-1910 - 1-Jan-1981</i>		<i>1-Jan-1900</i>
<i>GENDER</i>	<i>Numeric 1 (Nominal)</i>	<i>Sex</i>	<i>1 - 2</i>	<i>1 = female 2 = male 9 = missing</i>	<i>9</i>
<i>MSTATUS</i>	<i>Numeric 1 (Nominal)</i>	<i>Marital Status</i>		<i>1 = Single (never married) 2 = Married or living as married 3 = Widowed 4 = Divorced 5 = Separated 9 = Missing</i>	<i>9</i>
<i>FEET</i>	<i>Numeric 1 (Scale/Ordinal)</i>	<i>Height (ft)</i>	<i>4 - 6</i>		<i>9</i>
<i>INCHES</i>	<i>Numeric 2 (Scale/Ordinal)</i>	<i>Height (ins)</i>	<i>0 - 11</i>		<i>99</i>
<i>WEIGHT</i>	<i>Numeric 5.1 (Scale)</i>	<i>Weight (kg)</i>	<i>25 - 100</i>		<i>999.9</i>
<i>LSILL</i>	<i>Numeric 1 (Nominal)</i>	<i>Long standing illness</i>	<i>1 - 3</i>	<i>1 = Yes 2 = No 3 = Not sure 9 = Missing</i>	<i>9</i>

Variable Name	Type (Measurement)	Variable label (Recommended max of 40 characters)	Range	Value Labels (Recommended max of 20 characters)	Missing and N/A
<i>REGDISAB</i>	<i>Numeric 1 (Nominal)</i>	<i>Registered Disabled</i>	<i>1 - 2</i>	<i>1 = Yes 2 = No 8 = Not Applicable 9 = Missing</i>	<i>9 8</i>
<i>NOEX1</i>	<i>Numeric 2 (Nominal)</i>	<i>Reason 1 for no exercise</i>	<i>1 - 9</i>	<i>1 = Lack of leisure time 2 = Lack of money 3 = Lack of transport 4 = Lack of facilities at work 5 = Lack of interest 6 = Illness or disability 7 = Lack of Incentive 8 = Don't believe in exercise 9 = Other reasons 99 = Missing</i>	<i>99</i>
<i>NOEX2</i>	<i>Numeric 2 (Nominal)</i>	<i>Reason 2 for no exercise</i>	<i>1 - 9</i>	<i>* as for Reason 1 plus 88 = Not Applicable</i>	<i>88 99</i>
<i>NOEX3</i>	<i>Numeric 2 (Nominal)</i>	<i>Reason 3 for no exercise</i>	<i>1 - 9</i>	<i>* as for Reason 2</i>	<i>88 99</i>

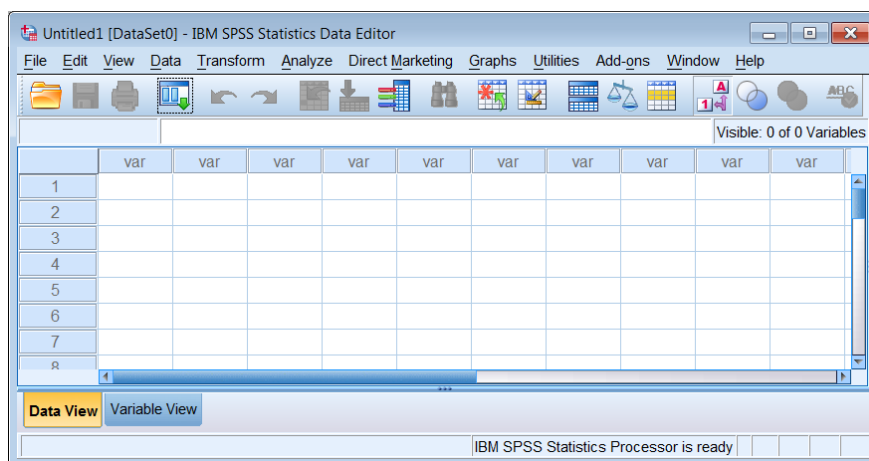
2 Running SPSS for Windows

To run SPSS either Double Click the SPSS button if it has been placed on the desktop.



Or Select  All Programs  IBM SPSS Statistics
 IBM SPSS Statistics 20 License Author
 IBM SPSS Statistics 20

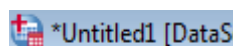
A panel asking you ‘What would you like to do’ is displayed. Select **Type in Data**. Click **OK**
An **Untitled1 [DataSet0] – IBM SPSS Statistics Data Editor** window is opened.



This window is like a spreadsheet workbook, but there are only two sheets. **Data View** is used for entering, editing and modifying data; and **Variable View**, is used to define the type of information that is entered in to each column in **Data View**. The Data Editor window is the main window used by SPSS, closing it ends the SPSS session. When you instruct SPSS to perform analyses or create charts, other windows are opened.

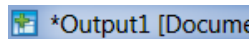
SPSS windows

Data Editor



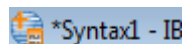
Is described above. You can have several Data Sets (Editor windows) open in SPSS.

Output Viewer



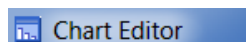
The results of statistical analyses and charts appear in this window. All printable output from SPSS is displayed here. The Viewer window is opened automatically when SPSS creates the first piece of output. The information displayed in this window (text, charts and pivot tables) can be edited and saved.

Syntax Window



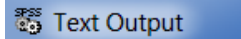
The SPSS system can be navigated by pointing and clicking at options on menus. However, it can also be controlled through a command language called SPSS syntax. The syntax window is where a SPSS program of commands is built, either by pasting commands from menus, or by directly typing commands into this window.

Chart Editor



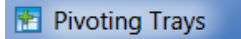
This is used to modify high resolution charts or plots. The editor can be invoked from the Output Viewer. Alternatively it can be used to edit a chart previously saved as a file.

Text Output Editor



This is used to edit text and change fonts of output in the Viewer.

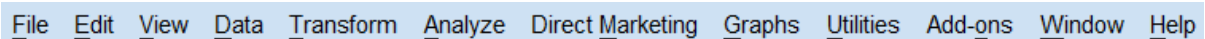
Pivot Table Editor



SPSS presents most of its output as Pivot Tables. This editor is used to modify Pivot Tables, for example, by swapping rows and columns. A Pivoting Trays window may be displayed depending on how you choose to edit your table.

SPSS menus

Each window in SPSS has a menu bar with specific options. The Data Editor menu bar has the following options:



- File** Is similar to most Windows applications. It provides the options for opening and saving files, displaying information about the current or another SPSS data file, printing the contents of the current window and exiting SPSS.
- Edit** Provides the usual undo, cut, copy and paste features, commands for finding and, with some windows, replacing text. SPSS programme settings can be changed using Options on this menu.
- View** Provides user control for the display of toolbars, the status bar, grid lines fonts and the display of values, or their labels, in the Data Editor.
- Data** Provides features to manipulate the data. Options in the Data menu can be used to define variables, copy the properties one variable to another variable, add new variables or cases, transpose variables, restructure the data set, select cases or subsets of cases for analysis and merge data files.
- Transform** The creation of new variables and recoding of existing variables are the major features offered by the Transform option.
- Analyze** This is the heart of SPSS. Options on this menu provide access to the statistical procedures for analysing data. All the items on the Analyze menu have sub menus.
- Direct Marketing** Provides a set of tools designed to improve the results of direct marketing campaigns by identifying demographic, purchasing, and other characteristics that define various groups of consumers and targeting specific groups to maximize positive response rates.
- Graphs** Provides options to create high quality plots and charts. An interactive chart builder is provided with legacy tools for previous graphical tools from SPSS.
- Utilities** Is used to display information on individual variables, add comments to the data file, define sets of variables to be processed and provides a spelling checker which is initially set to US-English.
- Add-ons** Provides access to Statistical Guides and information about other SPSS software products.
- Window** This allows all SPSS windows to be minimised. It also lists the currently open SPSS windows so that you can switch to the desired window; although this can also be done using the Windows task bar.
- Help** This is a typical Windows help menu which will be described later.

SPSS toolbars

Each window in SPSS has its own toolbar that contains buttons as shortcuts for frequently used menu options. Some buttons are common to all toolbars; others are specific to a particular window. The toolbar below is taken from the Data View of the SPSS Data Editor.

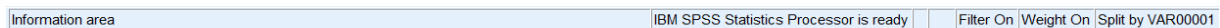


To find out what a button does, move the cursor over a button, after about one second a brief explanation is displayed.

The toolbars are customisable, you can add extra buttons to them or remove buttons that you use infrequently; you can also create your own toolbar. Toolbars can also be repositioned by dragging to the left, right, or bottom of the screen; or, they can become a floating window, if dragged anywhere else, inside or outside of the SPSS window.

Individual buttons will be described where they are used in these notes.

The Status Bar



The Status bar is positioned at the bottom of the each SPSS window. The information displayed on the status bar varies according to the window type. It may include:

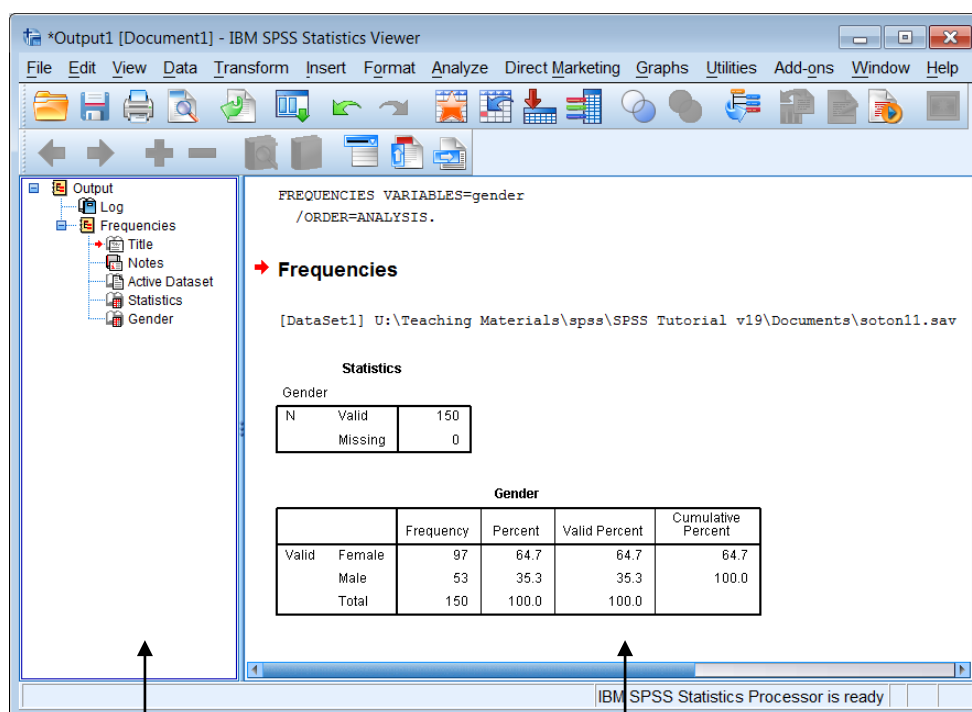
Information area	This is where explanations of buttons and other areas of the screen are displayed.
Command Status	This indicates the current status of the SPSS processor, e.g., ' <i>Starting SPSS processor</i> ' when the program loads, or ' <i>SPSS Statistics Processor is ready</i> ' when SPSS is waiting for instructions. When SPSS is executing a statistical procedure, the number of cases processed (so far) is displayed in this box.
Transformation pending	This indicates that SPSS is aware of some recoding or computation commands but has not performed the calculations yet. If a statistical procedure is requested these transformations are processed.
Filter Status	' <i>Filter On</i> ' indicates that some type of case selection or filtering is in effect for the current analysis and that only a subset of the data will be analysed.
Weight On	' <i>Weight On</i> ' indicates that each case is given different weights (by simulated replication) for statistical analysis. It allows summary information as well as raw data to be analysed by SPSS.
Split by Var....	' <i>Split by Var00001</i> ' indicates that the data set will be analysed by groups according to the variable indicated. VAR00001 in the example above.

The SPSS Viewer

SPSS displays its results in the Viewer. Usually it is opened when SPSS has the piece of output (e.g. results of a statistical procedure) to display. Publication quality output is produced by most procedures and all output is displayed in this window. The Viewer allows you to:

- Browse your results.
- Show or hide selected tables and charts.
- Change the display order of output by moving selected items.
- Edit pivot tables, charts, and text output.
- Move items between SPSS and other applications.

The Viewer is divided into two panes:



Outline

Display

The left pane of the Viewer, *Outline*, contains an outline view of the output contents. It is an index to the output. The right pane, *Display*, contains statistical output, charts, pivot tables and SPSS commands.

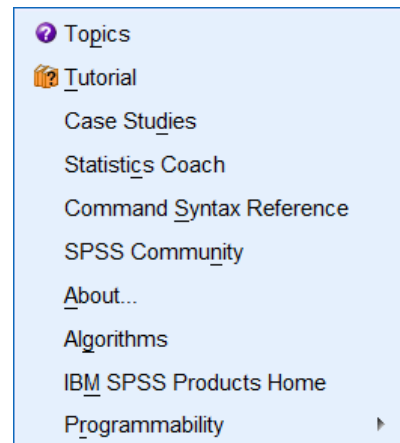
The outline pane provides a table of contents of all your output in a session. You can use the outline pane to navigate through your output and control the output display. Most actions in the outline pane have a corresponding effect on the Display pane as described on the next page.

- Selecting an item in the Outline pane displays the corresponding item in the Display pane.
- Moving an item in the Outline pane moves the corresponding item in the Display pane.
- Collapsing the outline view hides the output from all items in the collapsed levels.

Note: If one procedure produces a large amount of output, for example, a very large table, it may not be completely displayed in the Display pane. To view the contents the table must be viewed using the Pivot Table Editor.

Using the SPSS Help System

SPSS has extensive on-line help that is uses the standard Windows help system. Select Help from the menu bar by clicking on it and select one of the options.



Topics	This provides access to Contents, Index and Find as in all Windows help systems.
Tutorial	This provides step by step instructions for using SPSS. These relate to an earlier version of SPSS so the menu items are slightly different.
Case Studies	This is similar to the Tutorial option; it demonstrates how to perform certain operations using earlier versions of SPSS.
Statistics Coach	This assists in data analysis by leading you through a series of questions about your data and what you want to do with it.
Command Syntax Reference	SPSS, despite the menus, is a command drive program that uses commands that are described in the Command Syntax Reference which is a large PDF.
SPSS Community	Links to web site with resources for all levels of IBM SPSS product users and application developers
About	This provides details about the version of SPSS you are using and a link to the home page of SPSS.
Algorithms	Provides links to PDF files describing the mathematics of statistical methods
IBM SPSS Products Home	Links to www-01.ibm.com/software/analytics/spss/ . The SPSS site can be searched for keywords and is a very comprehensive source of help.
Programmability	This provides help on Scripting and use of the Python add in.

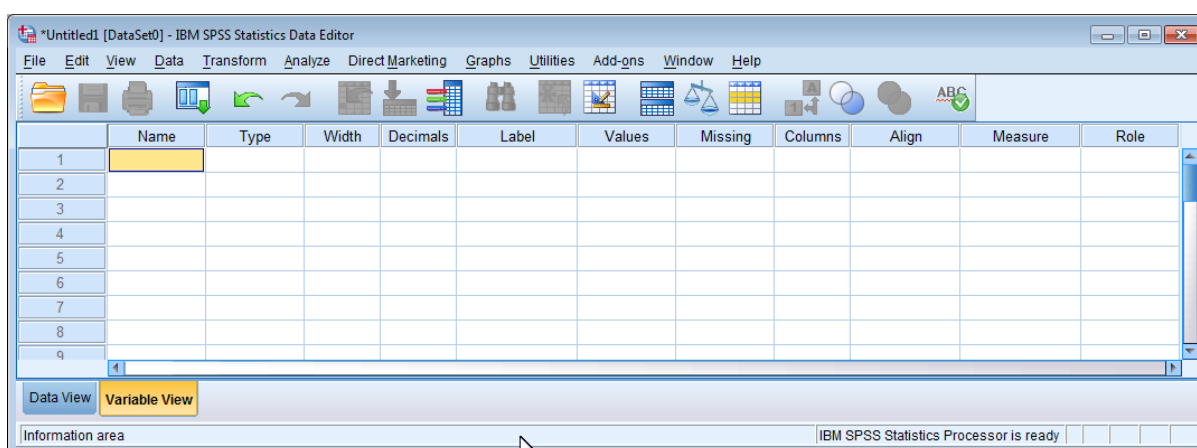
3 Defining variables using the Data Editor

The Data Editor acts like a spreadsheet, but there are some differences. The data file is rectangular consisting of rows and columns. Rows are cases and columns are variables. A cell is an intersection of a row and column and contains a data value. There are no unused cells in the data. Cells without values are considered to have a *system missing* value which is displayed as a full stop . for numeric variables and a blank (empty cell) for text.

There are two ways to define a variable, if the *Data View* sheet is displayed position the cursor over the first *var* box. **Double click** the left mouse button.



The alternative way is to ensure that the *Variable View* sheet has been selected, by clicking on it, as shown below.



The next section describes the information you need to be aware in order to set up variable information correctly. It is then followed by a step by step guide (see page 14) describing how to set up variables in SPSS. on how to do this using the coding guide data.

4 SPSS rules for defining variables

Variables Names

- Variable names must begin with a letter and can have a maximum of 64 characters. Names must be unique. You are recommended to keep names brief.
- The second and subsequent characters of any name can be a letter or a digit. You cannot use spaces in a name. The underscore character _ can be used where you want to use a space, e.g. REASON_1, REASON_2.
- Some symbols can be used but this is not recommended.
- Names are not cases sensitive so *GENDER*, *Gender* and *Gender* are identical. But the use of case is preserved for display purposes.
- Names that clash with Certain SPSS keywords are not allowed, e.g. ALL, WITH. SPSS will object if you try to use a reserved word.

Type of variable

SPSS assumes that all new variables are numeric. Click on the button in the Type cell to change the type of variable. The **Variable Type** window presents different option boxes depending on the type of variable currently selected. Numeric, as shown below offers width and decimal places. The three most used Types are:

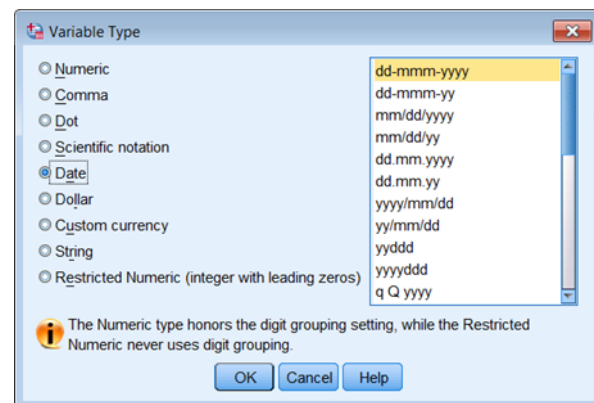
Numeric Width specifies the number of characters to be shown, this includes leading plus and minus signs (+ -) and the decimal point. So, if the maximum number encountered was -22.4, a setting of Width = 5 and Decimal Places = 1 would be appropriate. Width and decimal places are used for display purposes only and do not control the precision to which values are stored by SPSS.



Date Allows date and time formats, including days of the week and months of the year.

Use **dd-mmm-yyyy** for all dates.

The list also offers **hh:mm:ss**.



String Any letter, digit or character can be entered. SPSS has two type of strings:

- **Long Strings** have more than eight characters and could represent information, such as Name or Address (assuming Data Protection has been complied with).
- **Short strings** (8 or less characters). These can be used by various SPSS analysis procedures where are Long strings cannot.

Width

This specifies how many digits, for a number (maximum 40), or characters (maximum 32,767), for a string, are to be displayed. For decimal numbers this must include the dot. For example the value 23.1 has a width of 4.

Decimals


Specifies the number of values to be displayed after the decimal point (maximum 16). The value cannot be greater than the Width value minus one. This is sometimes a problem. If you want a variable to have a width of 1 (single digit) the Decimals value must be changed to zero before the Width can be changed to one.

Note The width and decimals are display and print options, they do not control internal precision. If the value 9.6789 were entered and the width were specified as 4 with 2 decimals, the value would be shown as 9.68, but SPSS would still retain the precision of the 4 decimal values internally.

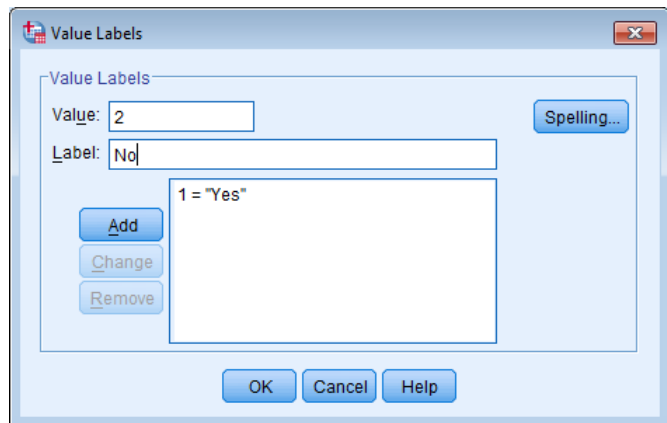
Label and Values

SPSS provides a label for each variable and the option to apply a label to each unique data value encountered for that variable.

- **Label**, also known as Variable Labels, can be up to 256 characters long (128 Chinese and other alphabets).
- **Value**, also known as Value Labels, can be up to 120 characters long (60 Chinese etc).

Labels and Values are displayed exactly as the characters are entered, case is important. Labels are typed into the cell directly, whereas Values are entered by selecting the cell, clicking on the button  and using the panel shown on the following page.

Labels cannot be specified for String variables that are greater than 8 characters long.



Adding a Value Label

Click in **Value** text box to enter a value.
The value may be numeric or text.
Click on the **Label** text box and enter a description.
Click on **Add** to add the label to the list.

Modifying a Value Label

Highlight the incorrect label in the list (bottom text box) by clicking on it.
Revise the label description in the **Label** text box.
Click on **Change**.

Deleting a label

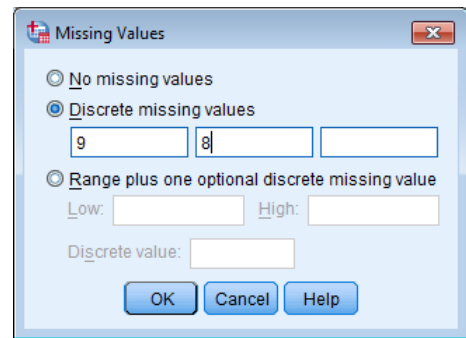
Highlight the **Label** in the list by clicking on it.
Click on **Remove**.

When you have added all the value labels, you can click **Spelling** to check for typing errors. Click **OK** to close the dialogue.

If you get a message '*Any Pending Add of Change operations will be lost*', this means you may not have added the last value label. You should click **Cancel** and complete your label.

Missing

SPSS provides two types of missing values:



System-missing values

Where no value has been entered. The cell contains a . (full stop or period) to indicate this.

User-missing values

Values which have a unique code(s) that you use to identify that the data was missing (e.g. 9). SPSS procedures can analyse data with User-missing values in several ways.

A valid date can be used to specify missing date values, for e.g., **01-Jan-1900** could be used.

The Missing Values options are:

No missing values

All values are valid.

Discrete missing values

Up to three different codes can be used to define missing values, for example, 8 and 9 might be used.

Range plus one optional discrete missing value


Numeric and Short strings (8 characters or less) can be used.

A Range of missing values and one optional single missing value can be used. For example 990 to 999 and -1.

Note: You cannot define missing values for strings (text) with more than 8 characters.

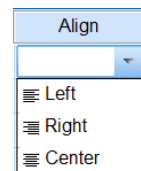
Columns

This specifies the width of the column as displayed in the Data View in the Data Editor.

Column widths can also be changed in the Data View by and dragging the column borders. Position the cursor over the right edge of the variable name cell. The cursor changes shape to a double headed arrow , drag the cursor horizontally to adjust the width of the column. (This is similar to changing column widths in Excel.)

Align

This specifies the alignment, Right, Centre or Left, of the data item.



Measure



Scale

Scale variables are numeric values which are measured on an interval or ratio scale for example, weight, or age.



Ordinal

Ordinal variables may be numeric or string categories that have a meaningful order; for example, Attitude scale categories where:

5 = Strongly Agree, 4 = Agree, 3 = No Opinion, 2 = Disagree, 1 = Strongly Disagree.

or Pain scale categories where :

a = No pain, b = Moderate pain, c = Severe pain and d = Unbearable pain.

We can use the value or letter to order the items, but the difference between 5 and 4 is not the same as the difference between 3 and 2, only that one is ordered (sorted) before the other.

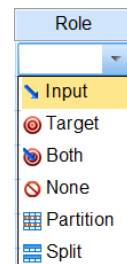


Nominal

Nominal variables can also be numeric or string, but the categories have no order; for example marital status – single, married, divorced, separated and widowed or blood group – A, B, AB, O.


Role


You can ignore this. Some SPSS dialogues support predefined roles that can be used to pre-select variables for analysis. When you open one of these dialogues, variables that meet the role requirements will be automatically displayed in the destination list(s).



5 Setting up variables in SPSS

The next section describes how to define the variable information. It uses a step by step approach. You might want to skip parts of the next few pages, but look at topics such as Copying an existing variable (page 18) which demonstrate time saving steps.

Items can be changed by clicking on the item to select the cell. Some items, for example, **Type**, **Values** and **Missing** will display a small box  in the right of the cell when selected. Clicking on this button will open an option dialog to do this. Other items, e.g. **Width** and

Decimal may display 'Spin Edit' arrows  that enable a value to be increased or decreased. However it is sometimes quicker to enter the required value using the keyboard.

Note: Some settings are inappropriate for some types of variable and are disabled.

To define a variable, position the cursor over the cell defined by **row 1** and column **Name**, and type the name of the variable, for example **id** and press either the <Enter>, or <Tab> key. Default values for the remaining columns, **Type** to **Measure**, are displayed by SPSS as shown below.

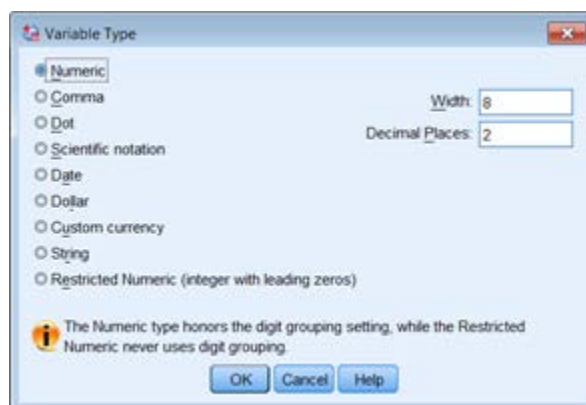
	Name
1	

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	id	Numeric	8	2		None	None	8	Right	Scale	Input

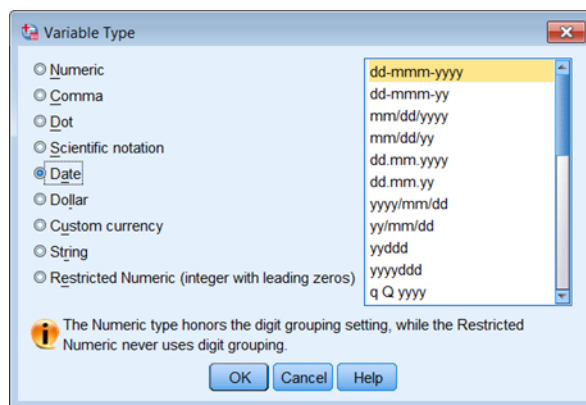
Click on the **Width** column and replace 8 with **4**. This can be done in one of two ways. Either press the 4 key or use the Spin Edit arrows to decrease the displayed value. Click on the **Decimals** column (or press Tab to move to it) replace 2 with **0**. Click on the **Label** column and enter **Case ID**. The information in the remaining columns does not need modifying.

Note: Variable Names are not case sensitive and can be entered as id, ID, Id or iD, they are always displayed as lower case by the Data Editor, e.g., id. Some parts of SPSS may display names in upper case (ID).

To define the next variable, **date of interview**, position the cursor over the cell defined by row 2 and column **Name**, type **doi**. Press <Enter> or click on the adjacent cell in the **Type** column. Click on the button in this cell to reveal the **Variable Type** dialogue box.



Click on **Date**, a drop down list of date types will appear.



Click on the **dd-mmm-yyyy** item.

Click on **OK**.

You will return to the Variable View sheet where the Type column row 2 will display Date. Width and Decimals will be displayed as 11 and 0 respectively in lighter grey (indicating that these two parameters cannot be modified).

Click on the **Label** column and type **Date of Interview**

In row 3 enter the next variable **DOB**, which is Question 1 on our survey form, by repeating the steps described above for DOI but using **DOB** for the variable name and **Date of Birth** for the Variable Label. Remember that the variable type is Date.

Now define the variable GENDER, question 2 in row 4.

Click on the *Name* column and enter **GENDER**

Leave the *Type* as numeric.

Click on the *Decimals* column, change the value 2 to **0**

Click on the *Width* column, change the value of 8 to **1**

Note: SPSS has an error check on the Width and Decimal columns, if you try to change Width to 1 with Decimal set at 2, which is what you wanted to do above, SPSS will object.

Click, or press <Tab> to move to the *Values* column. This is where you can add labels for the variable Gender where: 1 = Female, 2 = Male, 9 = Missing.

Click on the button  to reveal the *Values* panel.

Enter **1** in the *Value* text box

Click in the *Label* text box, enter **Female**

Click on **Add** (this button is highlighted as soon as a character is entered into the Value Label box)

Click in the *Value* text box in the Values Labels panel, enter **2**

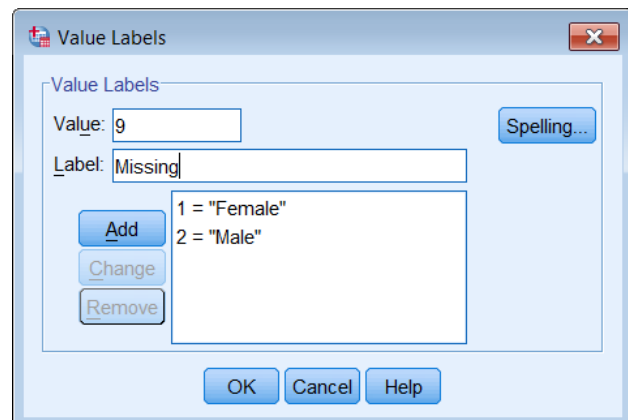
Click in the *Label* text box, enter **Male**

Click on **Add**.

Click in the *Value* text box in the Values Labels panel, enter **9**

Click in the *Label* text box, enter **Missing**

Click on **Add**.



You can **click** the **Spelling** button at this point to check your spelling of the labels.

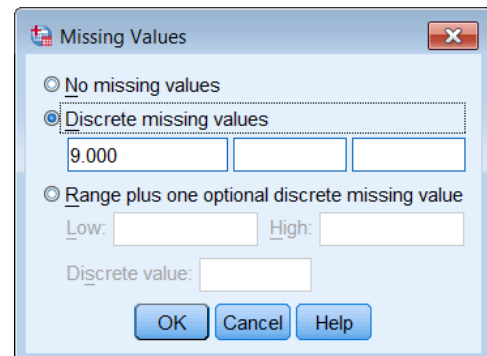
Click on **OK** to close the dialog

Click on, or press <Tab> to move to the *Missing* column.

Click on the button , the Missing Values dialogue box will be displayed.

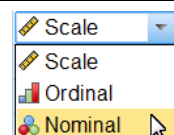
Click on *Discrete missing values*, enter **9** in the first box.

SPSS adds the three decimal places after you **Click** on **OK**.



Click, or press <Tab> to move to the *Measure* column.


Click the dropdown list and select **Nominal**.



Move to row 5 and enter **MSTATUS** in the *Name* column

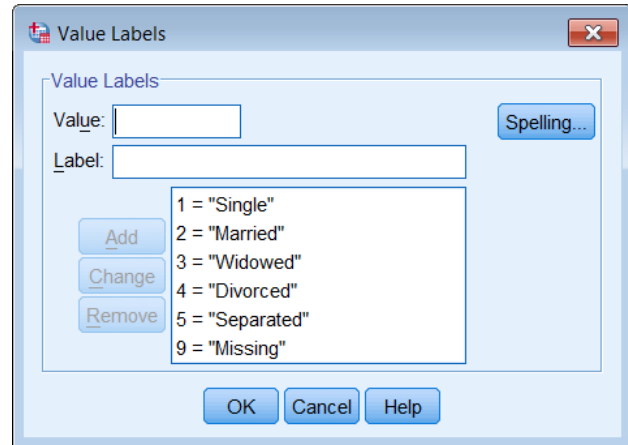
As for the Gender variable, leave the *Type* as numeric and change the *Width* to **1** and the *Decimals* to **0**.

Click on the *Labels* column, enter **Marital Status**
Click, or press <Tab> to move to the *Values* column.

Click on the button  to reveal the *Values* panel and enter the values and labels as for the Gender variable.

1 = Single
2 = Married
3 = Widowed
4 = Divorced
5 = Separated
9 = Missing

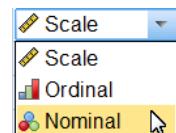
Click on **OK**.



Click, or press <Tab> to move to the *Missing* column.
Add a *Discrete missing values*, of **9**.

Click on **OK**.

Click, or press <Tab> to move to the *Measure* column.
Click the dropdown list and select **Nominal**.



The next question, question 4, asked for information about height, this needs two variables to be created, feet and inches.


Move to row 6

Enter **FEET** in the *Name* column.

Move past the *Width* column to the *Decimal* column and change 2 to **0**

Move back to the *Width* column and change 8 to **1**

Move to the *Label* column, type **Height - Feet**.

Move to the *Missing* column, **Click** on the button .

Enter **9** in the first *Discrete missing values* box, **Click** on **OK**.

To define the variable INCHES, repeat the same sequence as for variable FEET but use:
Name = **Inches**, Width = **2**, Decimals = **0**, Label = **Height - Inches** Missing = **99**

The question 5 variable, WEIGHT can be defined in a similar way, using:

Name = **Weight**, Width = **5** with Decimals = **1**, Label = **Weight - kg**, Missing = **999.9**

6 Saving the Data Editor Information

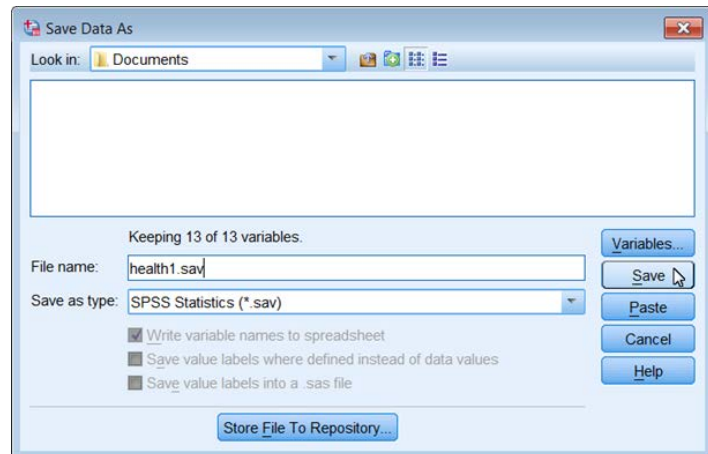
It is important to save your work at regular intervals, not just when the definition of your study is complete.

There are two ways to save the information

- 1) Click on the **Save File** button  in the SPSS toolbar
- or 2) Click on **File** on the SPSS menu bar and select either **Save** or, **Save As**

Either method displays the *Save Data As* dialog box .


Click on the *File name* text box and enter **health1**
Click on **Save**.



Note: the SPSS Data Editor window title changes from *Untitled* to **health1.sav [DataSet1]**

Question 6 asks about long-standing illness and disability.

Enter **LSILL** in the *Name* column, set Width = **1**, Decimals = **0**,
Label = **Long standing illness/disability/infirmary**


Move the cursor to the *Values* column, Click on the button 
Add in

- 1 = Yes,
- 2 = No,
- 3 = Not Sure,
- 9 = Missing

as shown

Click on **OK**.



Move the cursor to the **Missing** column, Click on the button 
Click on *Discrete missing values*, enter **9** in the first text box.
Click on **OK**.

Move to the *Measure* column, click on the drop box and select *Nominal*.

Copying an existing variable to create a new variable

The next variable REGDISAB is similar to LSILL so you can copy LSILL and make alterations to the copy to create REGDISAB.

Click on the next row number **9** . This highlights and selects the whole row.

9	lsill	Numeric	1	0	Long standing ... (1, Yes)...	9	8	Right	Nominal	Input
---	-------	---------	---	---	-------------------------------	---	---	-------	---------	-------

Click on **Edit | Copy** in the main SPSS window, or click the right mouse button and select Copy.

Click on the next row number **10**.

Click on **Edit | Paste**, or click the right mouse button and select Paste.

A new variable, called VAR00001 (or similar, e.g., VAR00002) will appear in row 10.

Change the variable name from VAR00001 to **REGDISAB**

Change the **Label** to **Registered Disabled**.

You need to replace one value label category, 'Not Sure', with another, 'Not Applicable'. Move to the Values column and **Click** on the button.

Click on the item **3 = "Not Sure"**, in the **Value Labels** dialogue box.

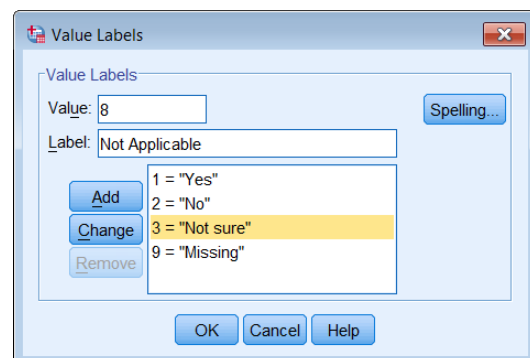
Click on **3** in the **Value** text box, change this to **8**


Click on **Not Sure** in the **Label** text box, change this to **Not Applicable**

(Both the Add and Change buttons will be enabled)

Click on **Change**.

Click on **OK**.



Move the **Missing** column, **Click** on the button 

Click on **Discrete missing values**

Add **8** in the second text box leaving 9 in the first text box.

Click on **OK**.

The next question in our data collection form asked why people did not take exercise. This is a multiple response question that is set up as three similar variables.

Set up the following information for row 11

Name = **noex1**, Decimal = **0**, Width = **2**, Label = **Reason 1 for no exercise**,

Missing = **88** and **99**, Measure = **Nominal**.

Highlight row **11** and copy it (**Edit | Copy** or click the right mouse button and select Copy), as you did previously for LSILL.

Highlight rows **12** and **13** and paste (**Edit | Paste** or click the right mouse button and select Paste).

Change the Names to **noex2** and **noex3**, change the Labels to **Reason 2 for no exercise** and **Reason 3 for no exercise**.

Add the labels for values for row 11, **noex1**, Reason 1 for no exercise.

Move to the **Values** column of row 11 (noex1), click on the button and create Labels.

Click in the **Value** text box, enter **1**

Click in the **Value** text box, enter **Lack of leisure time**

Click on **Add**.

Note: A variation using only the keyboard and not using the mouse is:

Click in the **Value** text box, enter **2** press the **Tab** key (cursor moves to the Value Label box)
 Enter **Lack of Money** press the **Tab** key (cursor moves to the Add button)
 Press **Space** bar (cursor moves to the Value box)

Enter the remaining labels.

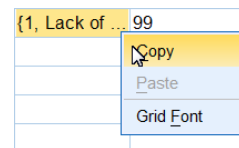
3	Lack of transport
4	Lack of facilities at work
5	Lack of interesting activities
6	Illness / disability
7	Lack of incentive
8	Don't believe in exercise
9	Other reason
88	Not Applicable
99	Missing

Do not forget to **Click** on **Add** for each label including the last label (99).
Click on **OK**.

Copying Variable Attributes

SPSS allows single attributes (e.g. Values) to be copied from one cell to another.

Move to the **Values** column of row 11 (*noex1*), **Right click** on the cell and select **Copy** to copy this cell's attributes (or click on the cell and use **Edit | Copy**).



Highlight the two label cells below it (**Click** and **Drag** down).
 Paste the attributes, **Right click | Paste** or **Edit | Paste**.

Note Any attribute from one variable can be copied to one or more cells.

The variable view panel should look similar to that shown below. The values in the **Columns** column may differ.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	id	Numeric	4	0	Case ID	None	None	8	Right	Scale	Input
2	doi	Date	11	0	Date of Interview	None	01-Jan-1900	11	Right	Scale	Input
3	dob	Date	11	0	Date of Birth	None	01-Jan-1900	11	Right	Scale	Input
4	gender	Numeric	1	0	Gender	{1, Female}...	9	8	Right	Scale	Input
5	mstatus	Numeric	1	0	Marital Status	{1, Single}...	9	8	Right	Nominal	Input
6	feet	Numeric	1	0	Height (feet)	None	9	8	Right	Ordinal	Input
7	inches	Numeric	2	0	Height (in)	None	99	8	Right	Ordinal	Input
8	weight	Numeric	5	1	Weight (kg)	None	999.9	8	Right	Scale	Input
9	lsill	Numeric	1	0	Long standing ...	{1, Yes}...	9	8	Right	Nominal	Input
10	regdisab	Numeric	1	0	Registered dis...	{1, Yes}...	9, 8	8	Right	Nominal	Input
11	noex1	Numeric	2	0	Reason 1 for n...	{1, Lack of ...	99	8	Right	Nominal	Input
12	noex2	Numeric	2	0	Reason 2 for n...	{1, Lack of ...	99, 88	8	Right	Nominal	Input
13	noex3	Numeric	2	0	Reason 3 for n...	{1, Lack of ...	99, 88	8	Right	Nominal	Input

You have finished defining the variables and so it is time to save the file again. Although you can Click on the Save button a better approach is to select **File | Save As** from the menu.

Save the file as **Health2**

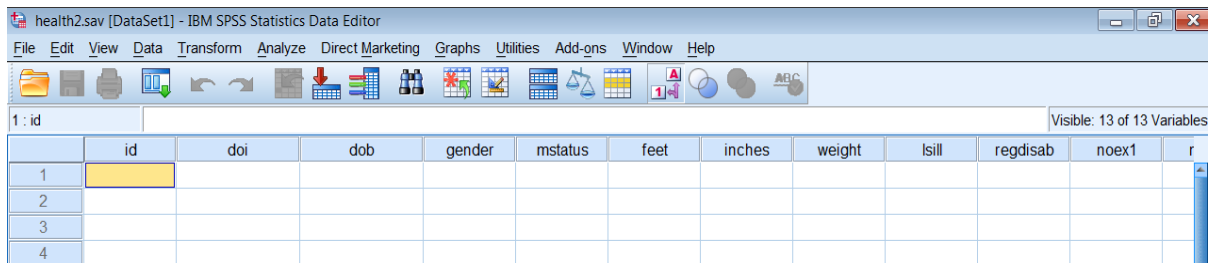
Note: Changing the file name is recommended. Files can occasionally become corrupted or overwritten accidentally. Using an incremental numbering system for file names allows you to keep the most recent files and delete older copies when they are no longer needed. If disaster strikes, you will only have lost perhaps half an hour's work, rather than several days. Make copies of your important files on other media, e.g. CD-ROM or temporarily on a USB memory stick.

Health2.sav contains the variable definitions for the study before any data is entered. A copy of this might be useful later.

7 Entering Data

Data can be entered into the Data Editor using techniques similar to those of a spreadsheet. Switch to the **Data View** sheet.

If you try to enter an invalid item into a cell, the Data Editor will change the cell to empty when you move to another cell.



Move the cursor to the first cell in the first row, the **id** column in the above example. **Click** on this cell and enter the data using the **<tab>** key or the **→** right cursor key to move across the row to add more data. If you press **<Enter>** key you move down the column.

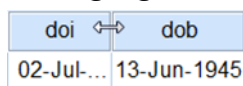
Dates can be entered in a various formats, for example 2.7.11, 2-7-11, or 2-Jul-11 are all valid ways of entering 2-Jul-2004. But 2/7/11 is not.

The century range for 2-digit year values is determined by your Options settings (Edit menu, Options, Data tab). The automatic setting is 1930 to 2029.

Data values for the first few (5) cases can be entered to check that you have set up the variable definitions correctly

26	02-Jul-2011	13-Jun-1952	1	2	5	8	999.9	2	8	99	88	88
29	08-Jun-2011	16-Apr-1975	2	2	5	11	999.9	2	8	1	99	88
41	19-Jun-2011	03-Jun-1947	1	3	5	2	68.5	3	8	2	6	7
59	14-Aug-2011	24-Feb-1994	2	1	6	4	54.0	2	8	99	88	88
69	03-Aug-2011	23-Mar-1982	1	1	5	4	77.0	2	8	1	2	5

Changing the width of a column



If the width of the column is too small, only part of the data value will be displayed. To increase the width of this column, move the cursor to the vertical line separating the columns for variables, e.g., **doi** and **dob**. The shape of the cursor will change to a double-headed arrow **↔**.

Drag the cursor to the right to increase the width of the column, **drag** it to the left to decrease the width. You can also change column widths by using the **Variable View** sheet and changing values in the **Columns** column.

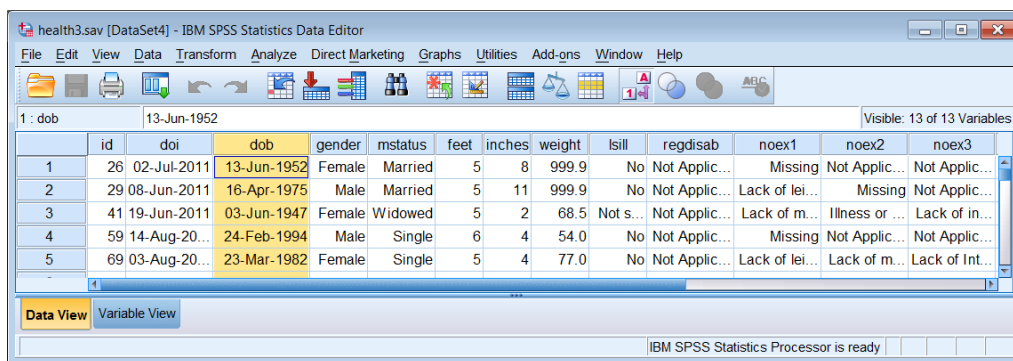
Problem you might encounter

- If you enter an invalid item, SPSS just ignores the value and leaves the cell contents as missing. Sadly it does not alert you in anyway.
- If a date of birth cannot be entered you may have forgotten to define the variable, as a date and left it as a numeric variable. Switch to the **Variable view**, – column **Type** and change the type to **Date**.

Key combinations for moving around the data editor

Action	Keys
Move down one cell	↓ or <Enter>
Move up one cell	↑
Move left one cell	← or <Shift Tab>
Move right one cell	→ or <Tab>
Select first cell (variable) in case (row)	<Ctrl ←> or <Home>
Select last cell (variable) in case (row)	<Ctrl →> or <End>
Select first cell in variable (column)	<Ctrl ↑>
Select last cell in variable	<Ctrl ↓>
Select first cell in data file	<Ctrl + Home>
Select last cell in data file	<Ctrl + End>
Scroll up or down the height of window	<PgUp> <PgDn>

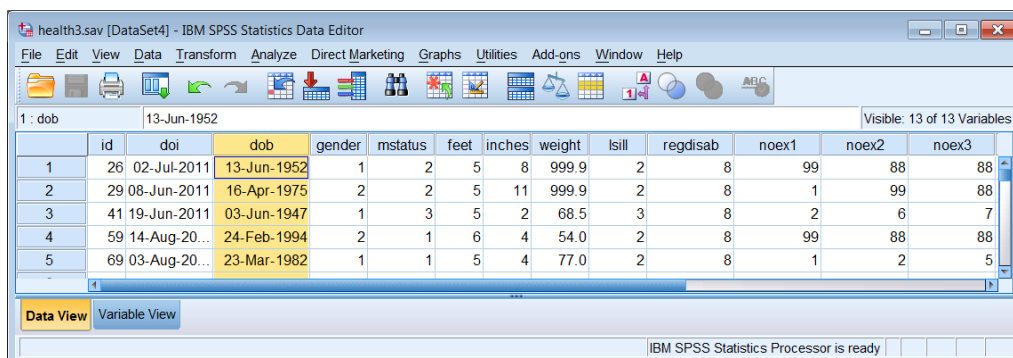
The screen below shows data for the first five cases.




Displaying Values or Value Labels

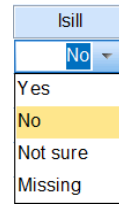
The data editor can display categorical information either as numeric values or their labels, if value label values have been defined. By default values are displayed, to change this either:

Click Values Labels button  on the Data Editor toolbar, or select **View | Value Labels**



To switch back to viewing numeric values, either **Click** on the  button again or select **View | Value Labels** again.

Note: All the possible labels for a variable can be viewed by clicking the drop down button.



Clicking on a label enters this value into the current cell. Numeric values can still be used for entering the data values when value labels are displayed (Even ones without labels!).

Correcting data

Replacing values

To replace an incorrect value.

Click on the cell whose value is to be changed and type the new value in to that cell.

Modifying values

To change part of a data value, for example a date:

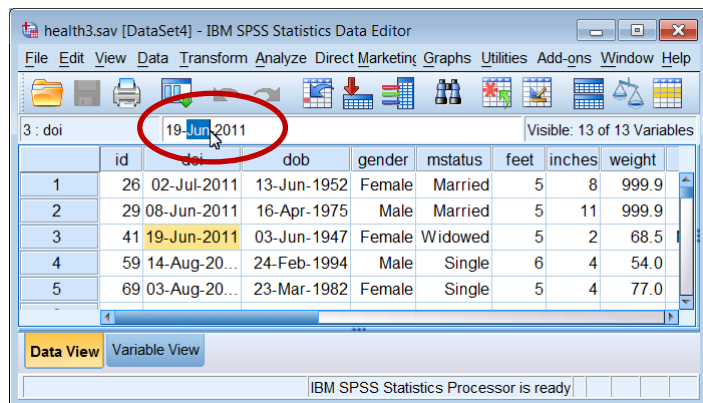
Click on the cell editor.

A flashing cursor (|) appears when you click on the cell editor. Edit data by using the mouse or the →←<Home> or <End> keys to position the cursor in the correct place.

To delete characters use <Delete> or <Backspace>.

To add characters type them.

Press <Enter> to complete the edit.



Other Data Editor Operations

Cut, Copy and Paste Data Values

These operations are performed in the **Data View** by:

Select the value(s) you want to Copy or Move, by **clicking** on a cell or **clicking and dragging** to highlight a group of cells.

From the Data Editor (Data View) window select **Edit | Copy** or **Edit | Cut**.

Select the target cell(s) by clicking on it (or first cell in a group of cells)

From the Data Editor (Data View) window select **Edit | Paste**.

Note : You can paste values of different data types as long as SPSS can convert them.

Add and Delete Cases

To insert a new case (row) into the data, **Click** on a cell **below** the position for the new case.

Either:

Click on the Insert Case button 

or Select **Data | Insert Case** from the Data Editor window menu.


New rows are created above the current row. All the values for the new case are set to system missing.

To delete a case (row), **click** on the number on the left side of the row, or press <**Shift Space**> with the cursor on that row. For multiple cases **Click and Drag** to highlight the rows.

Either press the <**Delete**> key, or from the Data Editor window select **Edit | Clear**.

Add and Delete Variables

To insert a new variable in *Data View*, **Click** on a cell. Either:

- Click** on the Insert Variable button 
- or Select **Data | Insert Variable** from the SPSS window menu.

The new variable (column) will be inserted to the left of the cell selected. (In Variable View it is inserted on the row above).

All cases are set to system missing for this new variable.

To delete an existing variable (column), **Click** on the variable name box. For several variables (columns) use the **Click and Drag** technique.


Either press the <**Delete**> key, or from the SPSS window select **Edit | Clear**.

Copying Variable settings

To copy the settings of a single variable and use them for another variable, select *Variable View*, **click** on a cell to select the variable item (e.g. Values). Select **Edit | Copy**. **Click** on the cell to receive these settings. Click on **Edit | Paste**.

If you only select the variable name, by clicking on the cell in the Name column, and use Copy and Paste. You will be warned that 'This variable name duplicates an existing variable'. The new variable will have default settings used by SPSS.

Change the order of (moving) variables

To change the order of variables, create a new variable in the position you want to move the existing one to by using  or **Edit | Insert Variable**. Then use a Cut and Paste technique to insert the variable into the correct location.

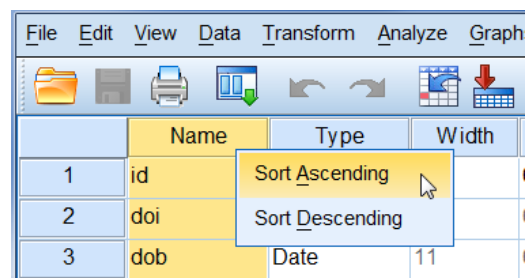
Delete the variable from its old location (see above), then rename the newly inserted variable from VAR0000? to that of the variable just deleted.

Note In *Variable View* only the variable information (variable type, label, value labels and missing values) will be copied. In *Data View*, both the data and variable information are copied.

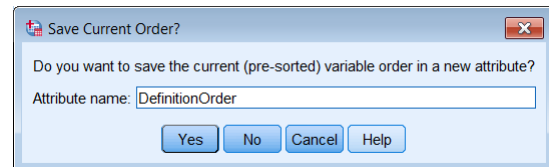
Sorting the Variable view

In Variable and Data View, variables are ordered by their creation order. You can order variables by any of the attributes you define for variables. For example variables could be ordered by their Name or their Measurement Type. To do this:

Move the cursor over a variable attribute column title, *Name* (as shown opposite) and **click the right mouse button** and select *Sort Ascending* or *Sort Descending*.



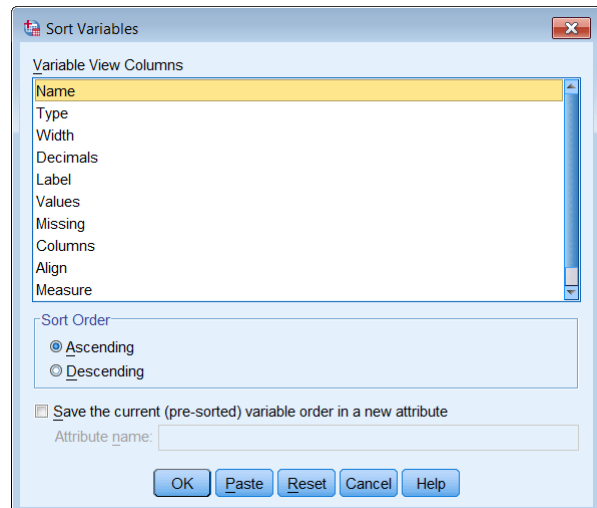
A *Save Current Order?* dialogue will be displayed. Enter an *Attribute Name*, e.g., **DefinitionOrder**. Click **Yes**.



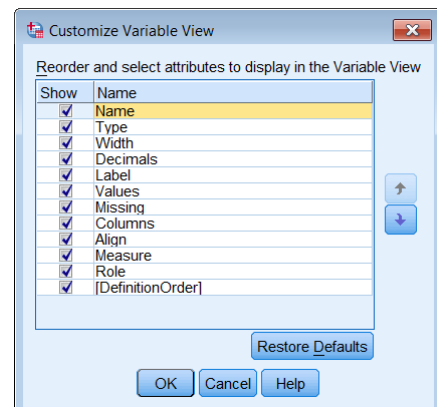
Alternatively you can select **Data | Sort Variables**. A *Sort Variable View* dialogue is displayed.

Select a variable from the *Variable View Columns*, choose the *Sort Order* and tick the *Save the current (pre-sorted)* box and enter an attribute name, e.g. **DefinitionOrder**.

Click **OK**.

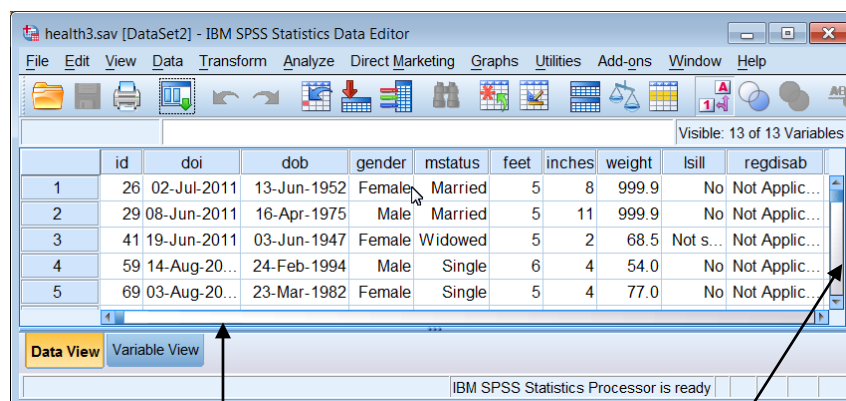


SPSS does not display the new attribute in the data editor, you must select **View | Customise Variable View** tick *[DefinitonOrder]* and click **OK**.



Multiple Data Views

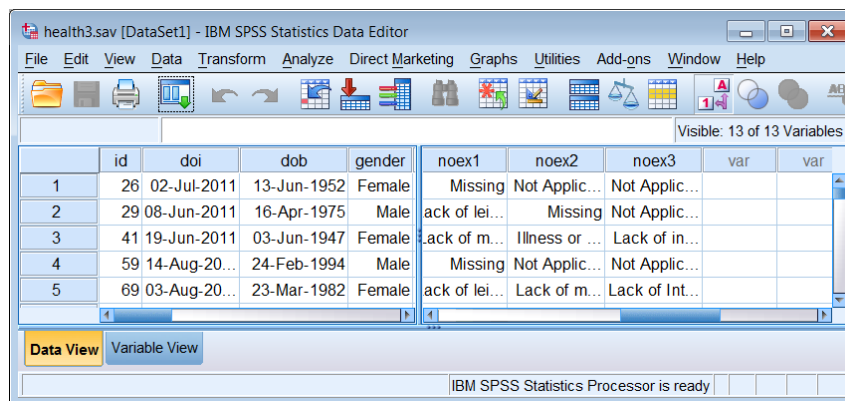
In *Data View*, you can create multiple views (panes) by using the splitters that are located below the horizontal scroll bar and to the right of the vertical scroll bar or selected from the menu using **Window | Split**. This process is similar to that found in Microsoft Excel.



Vertical Splitter

Horizontal Splitter

The splitters are below the horizontal scroll bar and to the right of the vertical scroll bar. The figure below shows the horizontal splitter dragged from its current position to mid point in the view. It is displayed as a thick blue vertical line during the dragging operation.



Use the scroll bars to view different columns of data as shown below:

Splitting is removed by selecting **Windows | Split** again.

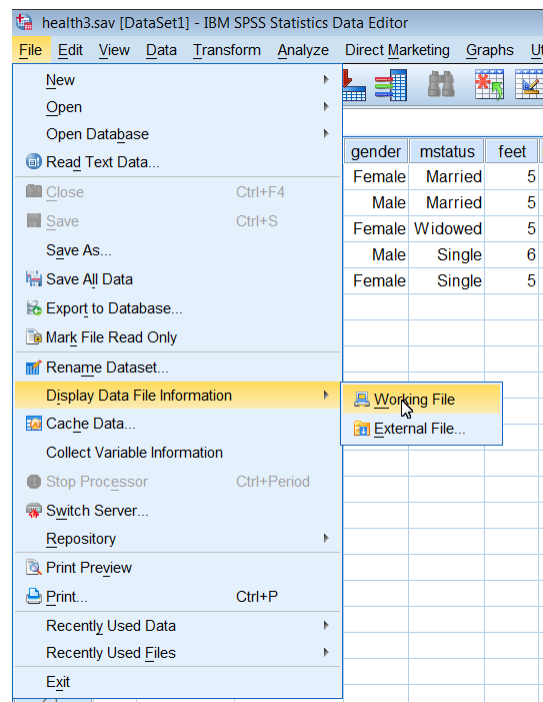
8 Checking Data

Before starting the statistical analysis it is important to check the data definition.

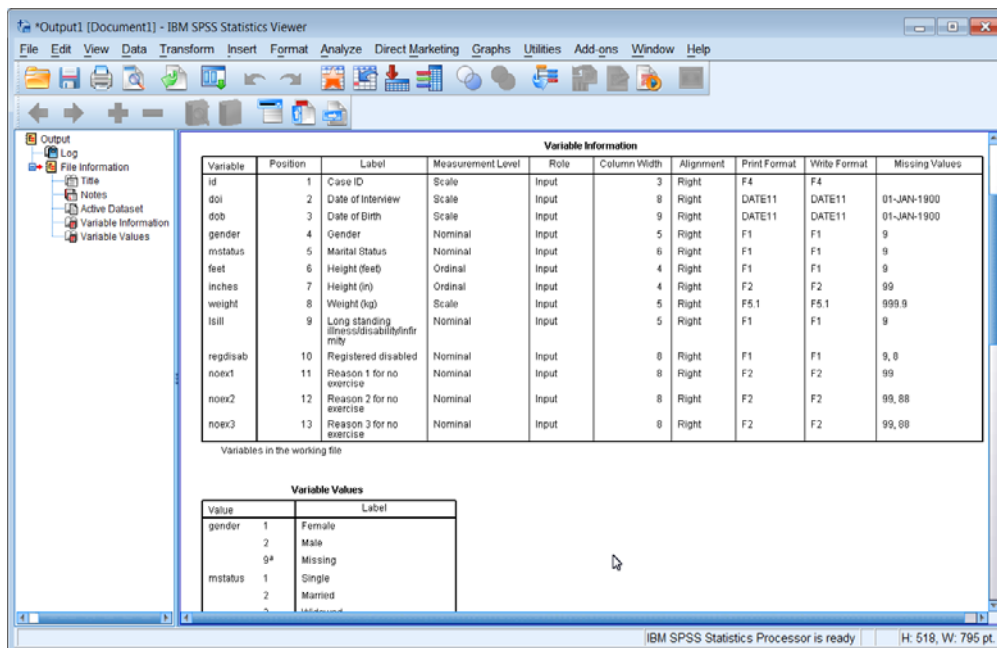
You can display and print the dictionary of the data file using **File | Display Data File Information | Working File**


SPSS will display all of the file information (dictionary) for each variable in the SPSS Viewer, as shown below.

Note The menus and toolbars in the SPSS viewer window have fewer options and buttons available. If these menus are not shown, select **View | Toolbars** and check the **Viewer Standard** and **Viewer Outlining Toolbars**.



9 Viewing Output



A useful tip is to remember to maximise the Viewer, if it is not already maximised. This makes it easier to view the output. Click on the maximise button  at the top right of the Output window. Use the scroll bars to view the all of the output.

Checking Variable Definitions

Details of the definitions of the first four variables are shown below. The output in the viewer can be used to verify that the variables have been defined correctly.

Variable	Position	Label	Measurement Level	Role	Column Width	Alignment	Print Format	Write Format	Missing Values
id	1	Case ID	Scale	Input	3	Right	F4	F4	
doi	2	Date of Interview	Scale	Input	8	Right	DATE11	DATE11	01-JAN-1900
dob	3	Date of Birth	Scale	Input	9	Right	DATE11	DATE11	01-JAN-1900
gender	4	Gender	Nominal	Input	5	Right	F1	F1	9


Value	Label
gender 1	Female
gender 2	Male
gender 9 ^a	Missing
mstatus 1	Single
mstatus 2	Married

The two tables Variable Information and Variable Values list the settings applied to these variables.

The next check you should do is to list the frequencies of all variables in the data set. This will allow you to check for errors in the data such as duplicate case numbers, or unexpected values. Although this can be done using menus, an easier way is to use the Syntax Editor window and type the command directly. Open a Syntax Editor window by selecting:

File | New | Syntax

or if a Syntax Window is already open,

Click on the Syntax Editor button  *Syntax1 - IB on the Windows task bar to make it the active window.


Type

freq all.

in the Syntax window.

Note: There is a full stop after the word 'all' which must be entered.

freq is short for Frequencies, the full word could be entered but it is not necessary.

Click on the **Selection** button  from the Syntax window toolbar to run this command.

Scroll through the output to check the frequency tables.

These frequency tables will help in the checking that:

- no duplicate case ids have been entered.
- values are within the expected range.
- missing values have been handled correctly.
- no unexpected values have been entered.

Further Checking

Further checking should be performed by cross tabulating categorical variables (for consistency checks and validation), producing histograms for continuous variables and scatter plots for pairs of continuous variables. These plots will give an idea of the average value, the

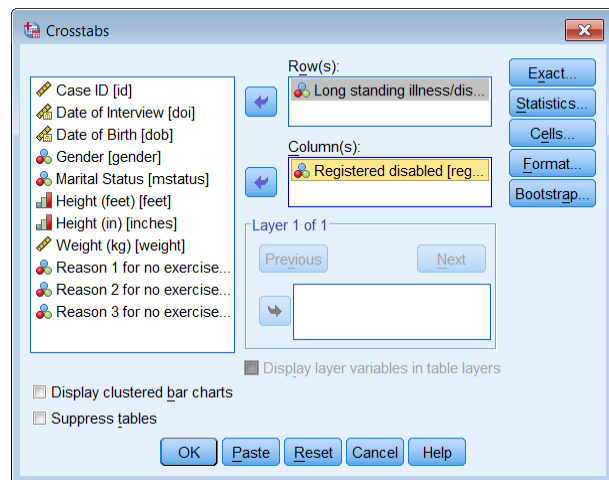
variability, the shape of the distribution and whether there are any outlying or missing values. This screening will help in the choice of statistical techniques used to analyse the data. Consistency of discrete variables (variables with a limited number of categories, e.g. gender) can be performed through the use of the Crosstabs command. For example, we may check the consistency of whether a case (i.e. respondent) with no long-standing illness has incorrectly selected "Yes" in response to the question, "If Yes, are you registered disabled"?

Select the Crosstabs dialogue box by clicking **Analyze | Descriptive Statistics | Crosstabs...**

Select *Long standing illness/disability/infirmity [lsill]* – by clicking on it, then click the  button to add it to the **Row(s)** panel.

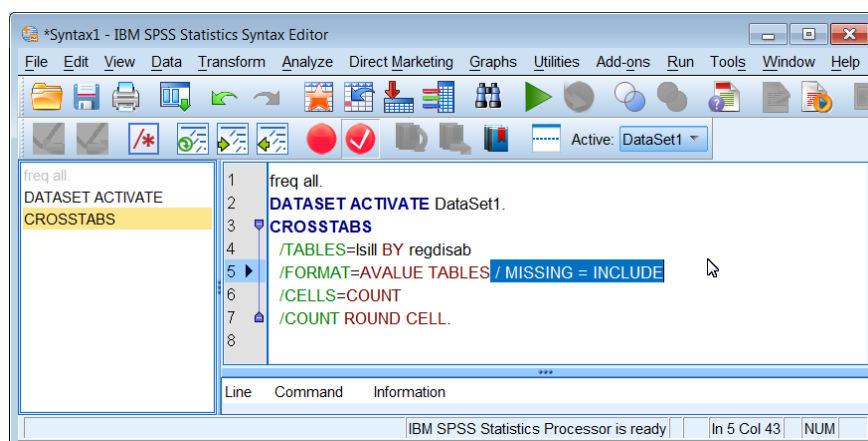
Select *Registered disabled [regdisab]* by clicking on it, click the  button to add it to the **Column(s)** panel. Click **Paste**.


Note An alternative method is to drag items from the list of variables into the Row and Column panels.




Switch to the *Syntax1 – SPSS Syntax Editor* window.

Add **/MISSING = INCLUDE** after the words **AVALUE TABLES** as shown below.



Run this Crosstabs command, by clicking the Selection command button . Without the **/MISSING = INCLUDE** added you would not get any output in this instance because there is no option in the CROSSTABS dialogue to include data identified as Missing.

10 Printing Output

Before printing any output it is a good idea to use the Print Preview which is similar to the preview feature in Word. This shows you what will print on each page. Ensure that the Output Viewer is the active window then click on the Print Preview button  or select **File | Print Preview**.

If this output is what you want you can click on the **Print** button in Print Preview.

You can choose what you want to print from the Output window, for example, specific tables or charts. Select by **Clicking** on items listed in the Outline pane. Use the Windows techniques of <Shift> or <Ctrl> keys to select multiple items.

Click on the printer button  or, select **File | Print**.

The **Print range** panel lets you specify either *All visible Output*, or the current *Selection*.

Click on **OK** to print.

*Note An alternative way to select items to print, is to delete items you don't want and then print all the output. To do this click on an item on the Outline pane and press **Delete** on the keyboard.*

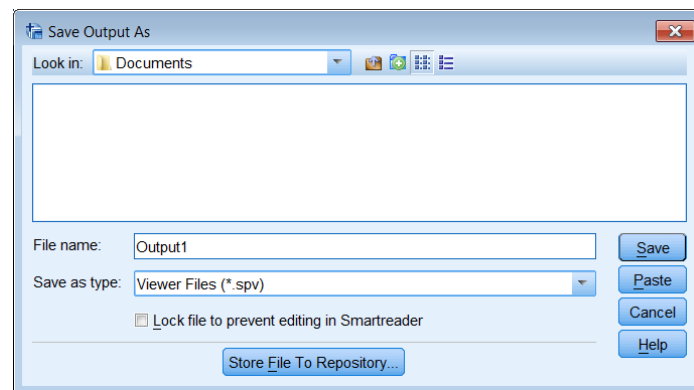
11 Saving Output

You can save the contents of the Viewer to a file but note that that Output files can occupy a lot of disc space.

The saved file includes both the outline and contents panes. The contents of the file can be reloaded into the Statistics Viewer at any time. Ensure that the Viewer window is the active window, by clicking on its button on the Windows task bar, or by choosing it from the **Windows** option on the menu bar.

Click on the Save button  or Select **File | Save As**.

Ensure that the folder you are saving to is MyDocuments and replace the file name with **Health1** in the **File name** box. **Click** the **Save** button. The file will be saved as Health1.spv.




From the Statistics Viewer menu you can use **File | Export** to export all or sections of the output. You can choose to do this with or without charts, or to export only charts (Graphics) in various formats including Excel, HTML, PDF, PowerPoint, Text and Word/RTF. Graphics can be exported as JPEG, BMP, PNG, EMF, TIFF or EPS formats.

12 Summarising Data

For the remainder of these notes we will use 150 cases from a data set similar to the one you set up earlier.

CLOSE SPSS before you start this section

- Go to web page www.som.soton.ac.uk/learn/resmethods/datasets/spss/
- Download the file **soton11.sav**, remember which folder it is saved to.
- The file will have a slightly different name, e.g. **soton11-sav-1261.sav**.
- You might want to copy the file to the Desktop and rename it to **soton11.sav**.so that the folders and filename used in these notes are identical.
- Start **SPSS** and cancel the opening dialogue.

Click on the Open file button .

Select the Desktop folder and select file *soton11.sav* using the Open Date File dialogue box.

Click on **Paste** (Not on OK).

This will enter two commands into the SPSS Syntax Window.

GET

FILE='C:\Users\tnb\Desktop\soton11.SAV'

DATASET NAME DataSet1 WINDOW=FRONT.

Note The folder name (C:Users\tnb), will reflect your username.

Highlight these two commands, either by:

Holding the left mouse button down and dragging over the text.

Or by Pressing <Ctrl A>

Or by Selecting **Edit | Select | All**.

Click on the **Selection** button .

The next few pages will demonstrate the use of some of the **Analyze** commands in SPSS.

Other sources of information are,:

- **Which Statistical Test?**
www.som.soton.ac.uk/learn/resmethods/statisticalnotes/which_test.htm
- **SPSS statistics coach** may assist in choosing the appropriate statistics, chart or statistical test for summarising your data. To use the statistics coach, from the menus in any SPSS window select: **Help | Statistics Coach**
It is designed to provide general assistance for many of the basic, commonly used statistical techniques.


Categorical (nominal, ordinal) data

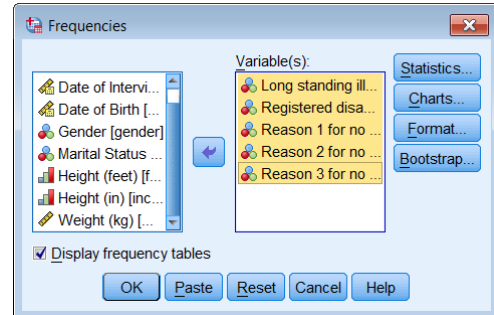
The **Frequencies** command is most often used for analysing categorical data. This command produces frequency tables that display the count of the number of times each code (category of the variable) occurs. Select:


Analyse | Descriptive Statistics | Frequencies

Highlight variables *lsill noex1 noex2 noex3 regdisab*


(Click on *lsill* then hold down the **Ctrl** key and click on each of the remaining variables).

Drag these items or click on the  button to move these variables to the **Variable(s):** panel.




Note: If you select the wrong variable and add it to this list. Highlight the wrong variable in the Variables list box and either drag it back to the variable list (on the left) or click the button  to remove it.

Click on **Paste**.

Select the **SPSS Syntax Editor** button on the Windows taskbar  *Syntax1 - IB this will show the following command has been pasted into the Syntax window:

```
FREQUENCIES VARIABLES=lsill regdisab noex1 noex2 noex3  
/ORDER ANALYSIS .
```

Click on  to run this command. (As this is a single SPSS command it does not need to be highlighted before clicking on the Run Current syntax button).

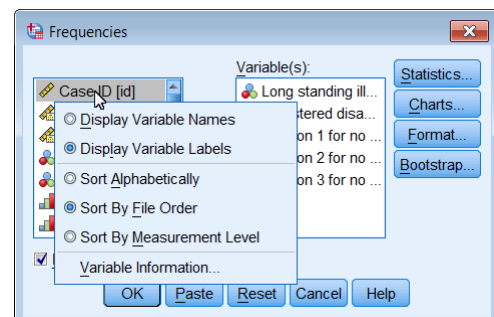
The results appear in the Statistics Viewer window. They can be viewed as explained earlier.

SPSS can show the list of variables in various ways.

Right Click

Continuous (scale, interval, ratio) data

The **Frequencies** command is just one of the commands that can be used to analyse variables of continuous data. An example is the weight of respondents, variable *weight*. Select the Frequencies dialogue box again either by:



Analyse | Descriptive Statistics | Frequencies

Or Click on the Dialogue Recall button . A drop down list with Frequencies on it will appear. Click on *Frequencies*.

Click on the **Reset** button.

Select *weight* and drag it to the Variable(s) panel.

Uncheck the *Display frequency tables* box.

An information message box about 'all output is turned off' will be displayed. **Click** on **OK**.

Click on the **Statistics** button.

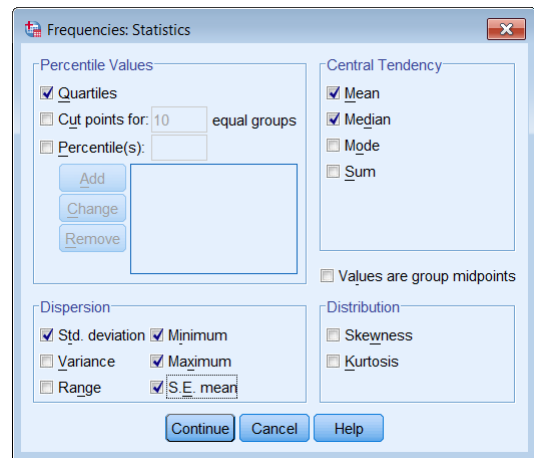
A *Frequencies: Statistics* dialogue box will be displayed.

To select the required statistics:

Check the *Quartiles* boxes in the *Percentile Values* panel.

Check the *Mean* and *Median* boxes in the *Central Tendency* panel

Check the *Std. deviation*, *Minimum*, *Maximum* and *S.E. mean* boxes in the *Dispersion* panel.



Click on **Continue**.

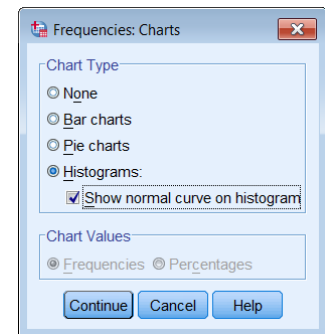
Click on **Charts** in the *Frequencies* dialogue box.

A *Frequencies: Charts* dialogue box will be displayed.

Check the *Histograms* button and the *With normal curve* box in the *Chart Type* panel.


Click on **Continue**.

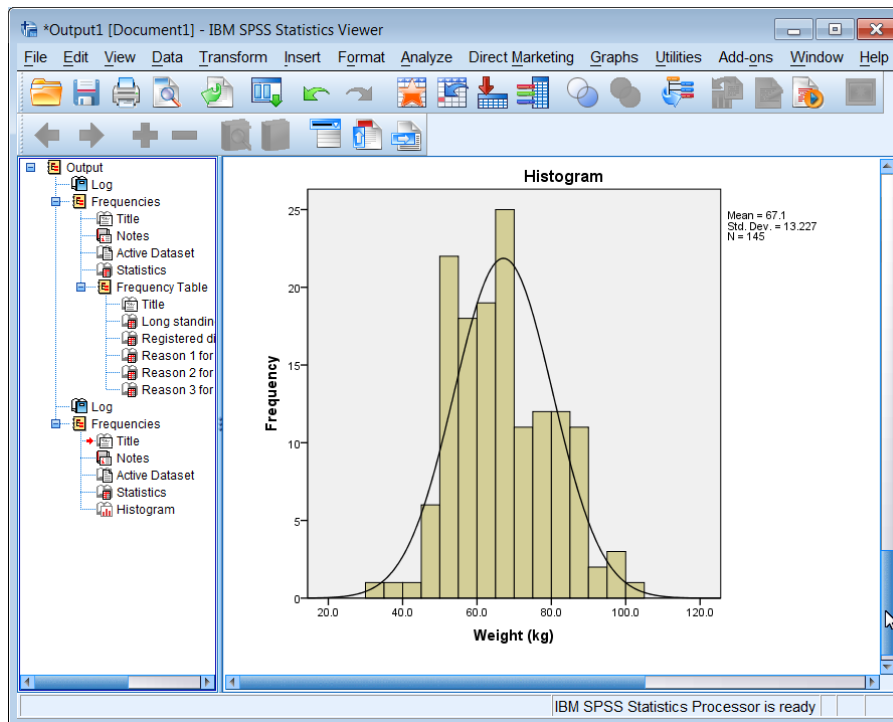
Click on **Paste**.



Move to the *Syntax1 - Syntax Editor* window to see that the following command has been pasted into the Syntax1:

```
FREQUENCIES VARIABLES=weight  
/FORMAT=NOTABLE  
/NTILES= 4  
/STATISTICS=STDDEV MINIMUM MAXIMUM SEMEAN MEAN  
MEDIAN  
/HISTOGRAM NORMAL  
/ORDER ANALYSIS .
```

Click on  to run the command.



View the results in *Output*, window. The outline pane lists the output, including the histograms that have been created. To view the histogram of weight, **Click** on Histogram listed in the Outline (left) panel. You may need to move, by clicking and dragging the right border of the outline pane to read clearly the outline list.

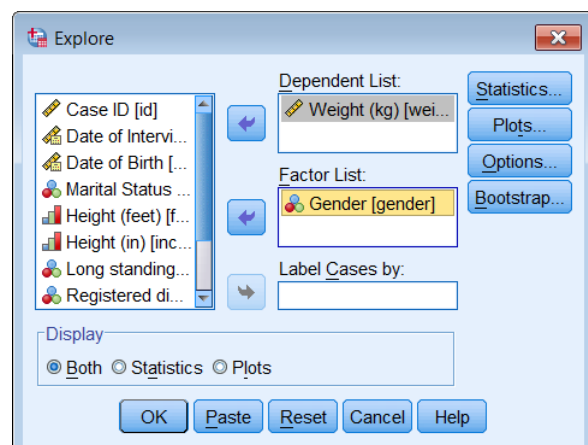
It is also possible to use **Analyze | Descriptive Statistics | Descriptives** to produce similar statistics but this procedure does offer the option to create charts.

Another very useful command used with continuous variables is the **Explore** command. It is similar to the **Frequencies** command but it possesses a greater variety of plots and it can analyse subsets (specified as factors) of the data set. For example, suppose we want to analyse the weight of respondents by Gender.

Select **Analyze | Descriptive Statistics | Explore**

Highlight *weight*, add it to the *Dependent List* panel.

Highlight *gender*, add it to the *Factor List* panel.



Note: If you do not specify a Factor List variable, Explore analyses all cases for each variable in the dependent list.

Click the Plots button.

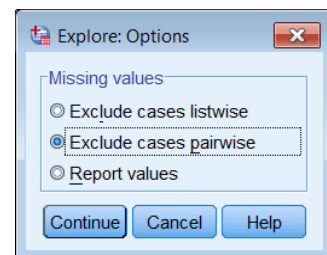
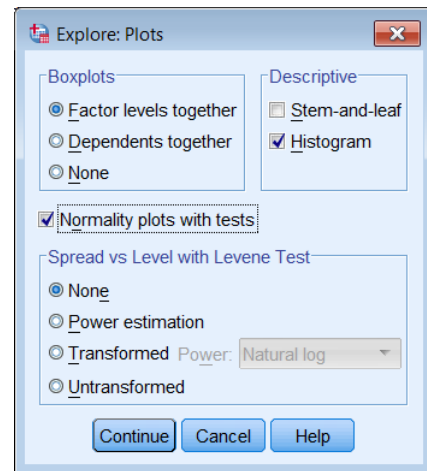
UnCheck the *Stem-and-Leaf* box on the *Descriptive* panel.

Check the *Histogram* and *Normality plots with tests* check boxes .

Click on **Continue**.

Click on **Options**.

Click the *Exclude cases pairwise* option button.



*Note: Although the Exclude cases pairwise option is not needed in this example, because we are only using two variables, this option is **important** when you are analysing several variables using the Explore procedure. The **Exclude cases pairwise** option uses the maximum information in a data set. Cases are included in the analysis even if they have missing data for some cases for some variables in this list. **Exclude cases listwise** excludes a case from analysis if data is missing for just one case for one variable.*

This option for the treatment of Missing Values is found in many SPSS procedures.

Click on **Continue**.

Click on **Paste**.

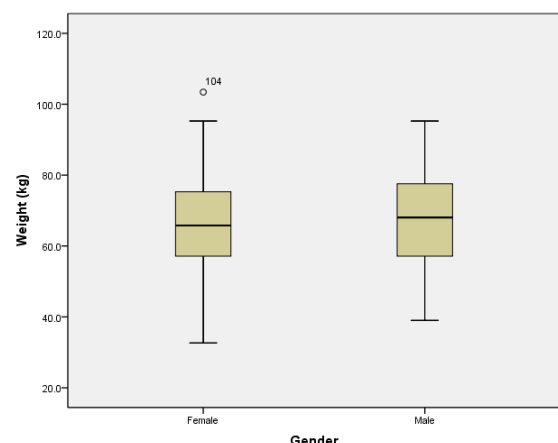
The Explore dialogue box creates an SPSS command called EXAMINE (not Explore!). It produces high resolution charts of Box plots and Histograms.

```

EXAMINE VARIABLES=weight BY gender
/PLOT BOXPLOT HISTOGRAM NPLOT
/COMPARE GROUPS
/STATISTICS DESCRIPTIVES
/CINTERVAL 95
/MISSING PAIRWISE
/NOTOTAL.
    
```

Click on  .

Note Graphs can also create the same chart.



Crosstabulations

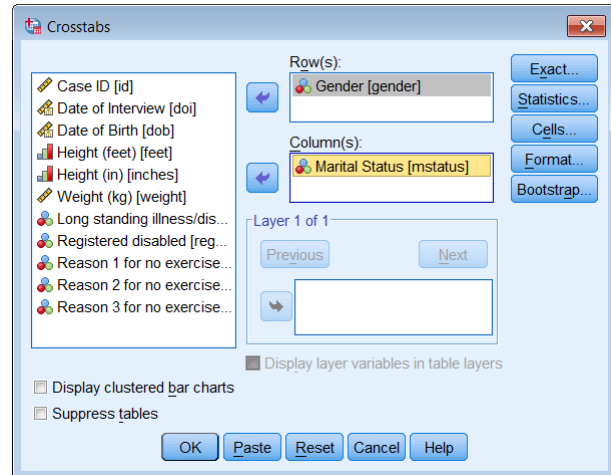
Tabulation of one categorical variable against one or more other categorical variables is done using the **Crosstabs** command. To cross tabulate *gender* by *marital status*, select:

Analyse | Descriptive Statistics | Crosstabs

Highlight *gender* and add it to the Row(s) list. (In the Output, row categories are printed down the page).

Highlight *mstatus* and add it to the Column(s) list.


In the Output, column categories will be printed across the page.

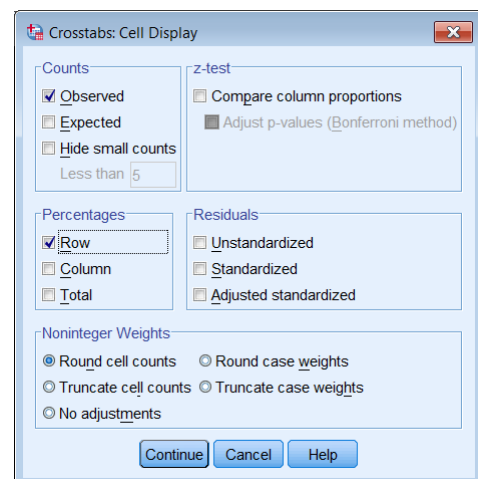


Click on **Cells**

Click on the **Row** check box in the **Percentages** panel.
Click on **Continue**.

Click on **Paste**.

Click on  .



Other buttons on the Crosstabs dialogue allow the selection of Statistics, for example Chi-square. At this stage the cross tabulation table should be examined and percentages compared before any statistical hypothesis testing is performed.

13 Why all this pasting? - the SPSS Syntax file

You have already used the syntax window to execute SPSS commands and by typing **freq all** into the window. Although SPSS uses menus and dialogue windows, it is basically a command driven program. The syntax window is where commands are pasted, or typed directly, to build a SPSS program. This program can be saved to a file. It can be recalled and executed again at a later date if the data changes. The importance of entering (pasting) commands into the Syntax window and saving them cannot be over emphasised.


Note: There are some SPSS commands and Options for some SPSS commands that can only be entered using the Syntax window because dialogue boxes for building them do not exist.

Saving and Printing the Syntax window

To save the contents of the syntax window ensure that it is the current window (The title of this window must be **SPSS Syntax Editor**. Click on the button on the windows task bar, or select it from the Window menu).


Either Select

File | Save As or **File | Save**

or **Click** on the Save button .


Chose an appropriate Documents folder and enter **Health1** (your chosen name in this example) into the **File name** text box and ensure that the directory and drive destinations within this dialogue box are correct.

Click on **Save**.

Note The name of the syntax window changes to **health1**. SPSS automatically adds .sps as the file extension. From now on **File | Save** or clicking on  will save the contents of this window to health1.sps without asking if it should replace the existing contents.

Retrieving and Editing the Syntax file

To retrieve syntax commands that were previously saved.



Either select **File | Open** or Click on  on the Syntax window toolbar.

Note: If the current window is not a Syntax Window, the **Files of type** box must be scrolled and select **Syntax (*.sps)**.

The **Open** dialogue box will be displayed (it is identical to the Save dialogue box).

Select the appropriate drive, folder and filename details.

Double click on the filename or **Click** on **Open** and the file will appear in a second syntax window.

- Commands can be typed directly into the syntax window, or pasted from dialogue boxes.
- To execute one command at a time, ensure that the flashing insertion point is placed somewhere on the command line(s), and then click on .
- To run multiple commands, you must highlight all the lines containing the commands before clicking . To do this move to the top left or bottom right of the command lines, hold down the left mouse button and drag the mouse to the opposite corner


(Note: Only part of the first and last commands must be highlighted)


To execute all the lines within a syntax window.



Select **Edit | Select All**.

or Press **<Ctrl A>**

All the lines in the syntax file will now be highlighted.

Click  to execute all the commands.

When the Syntax window is active, clicking on the syntax button  will launch a 'Syntax Help' window with details of the syntax for the current command. A general syntax help window will be presented if the cursor is not on a command.

Tip: It is possible to have more than one Syntax window (file) open concurrently. However, when more than one Syntax window is open, commands are pasted into the active or Designated Syntax window. To make an alternative Syntax window the Designated window by click the  button. The active or designated syntax window shows a disabled designate  button on the toolbar

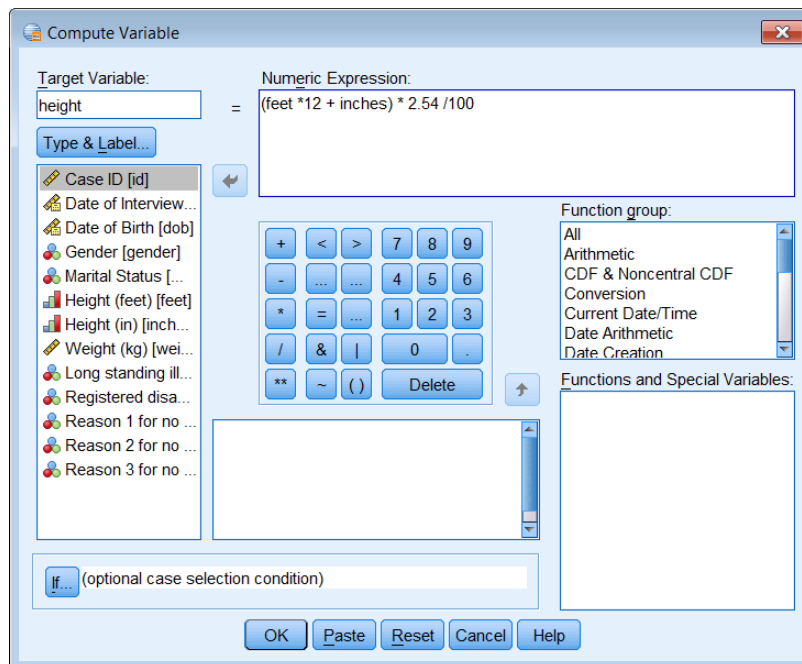
14 Creating New Variables and Modifying Data

In this section, features of SPSS that enable you to create new variables and modify existing variables will be described.

Suppose you wish to calculate the height of each respondent in metres and you have collected the data as two variables, feet and inches.

Go to the Data Editor window, Select **Transform | Compute Variable**.

The figure below shows the completed window for this.



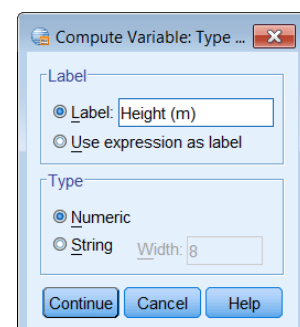
A new variable name, *height*, was typed into the *Target variable* box.

The *Numeric Expression* box is used to enter the formula used to calculate the formula for the target variable. The expression can be built either by typing it directly using the keyboard or it by clicking on the buttons on the calculator pad, the *Functions and Special Variables* list and the list of variables.

Items are added to an expression at the current cursor position, so you may need to move the cursor and click it to reset the insertion point in an expression. To delete an item, or sequence of characters, highlight the item/characters using the click and drag technique and then click on the **Delete** button on the calculator pad, or press the <Delete> key on the keyboard.

The formula for using the two variables, *feet* and *inches*, was entered as feet multiplied by 12 plus the value of inches, this total height in inches was then multiplied by 2.54 (inches to centimetres conversion) and then divided by 100 to obtain the result in metres.

Click on the **Type & Label** button to displays a dialogue box which the Variable *Label* information can be typed into.



The commands pasted into the Syntax Editor are:

```
COMPUTE height=(feet +inches*12) * 2.54/100.  
VARIABLE LABELS height 'Height (m)'.  
EXECUTE.
```

15 The Calculator Pad

The calculator pad has a number of arithmetic, relational and logical operators. Although these operators are presented as symbols, logical operators can be entered also as two or three character mnemonics. These are included in the table below.

Symbols and their meanings

+	Addition	<	Less than	LT	>	Greater than	GT
-	Subtraction	<=	Less than or equal to	LE	>=	Greater than or equal to	GE
*	Multiplication	=	Equal to	EQ	~=	Not equal to	NE
/	Division	&	Logical AND	AND		Logical OR	OR
**	Exponentiation	~	Logical NOT	NOT	()	Parenthesis	

Arithmetic operations have the following order of evaluation. Functions (e.g. LOG, MEAN) are evaluated first, followed by exponentiation (**), then multiplication / division and finally addition / subtraction. The order of calculation can be controlled by parentheses. Expressions inside parentheses “()” are evaluated first. So A+B/C means divide B by C and add the result to A, whereas (A+B)/C means add A and B together and divide the result by C.

The Logical and Relational operators are used when the creation of a variable is conditional on an expression. These are commonly used after the **If...** button has been clicked. This is described later.

Functions

SPSS for Windows provides many functions that can be used in expressions; these are grouped by function in the Compute dialogue window (*Function group*) and displayed in the *Function and Special Variables* list. They include:

- Arithmetic (ABS, RND, TRUNC, MOD, etc.)
- Statistical (SUM, MEAN, MIN, MAX, etc.)
- Search (RANGE and ANY, true if value is in range or list)
- Various Date and Time manipulations
- String and conversion functions

Select **Help | Command Syntax Reference**. Navigate the Command Syntax Reference to the *COMPUTE* section (page 332 -v18).

Missing Values and Functions

Functions treat missing values in a more useful way compared to simple arithmetic expressions. For example, if you wanted the mean of a set of variables, V1, V2, V3, using

```
COMPUTE X = MEAN(V1,V2,V3)
```

will calculate a mean for these three variables even if one, or more, of them has a missing value, whereas a statement such as

```
COMPUTE X = (V1+V2+V3)/3
```

will produce a missing result if any of the values of V1, V2 or V3 for a case are missing.

Rule for expressions

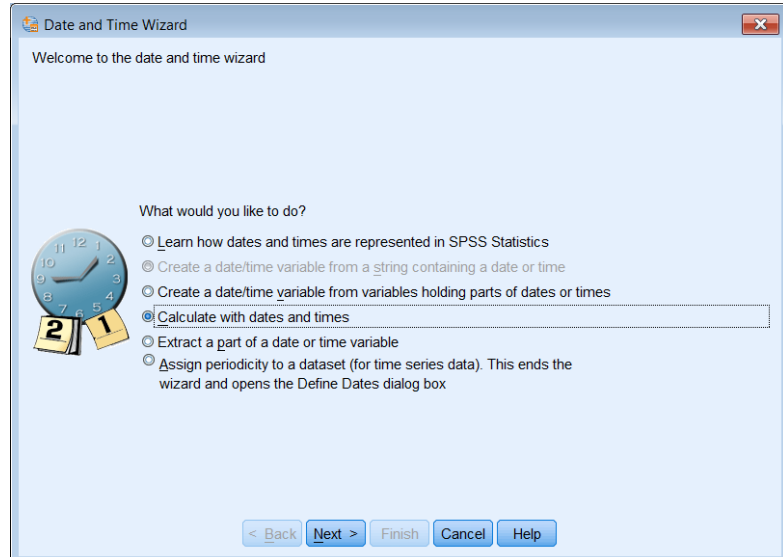
- The argument list for a function must be enclosed in parentheses.
- If a function has multiple arguments these must be separated by commas, spaces can be added for clarity.
- When comparing strings, text must be enclosed in either single ' or double " quotation marks, e.g. Sex = 'Male'. Otherwise Sex=Male would be trying to test the equality of two variables, Sex and Male.

The Date and Time Wizard

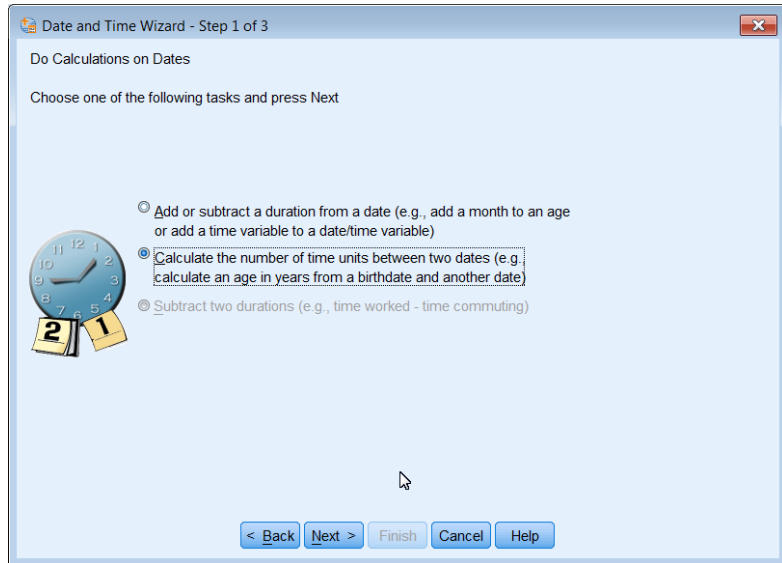
SPSS has a wizard to help you manipulate date or time variables and create new variables. It is selected from the menu, **Transform | Date and Time Wizard**. Choose options to help you generate a new variable. The sequence below shows how Age can be calculated from two dates.

First choose what type of calculation you wish to perform, here it *Calculate with dates and times*.

Click **Next**

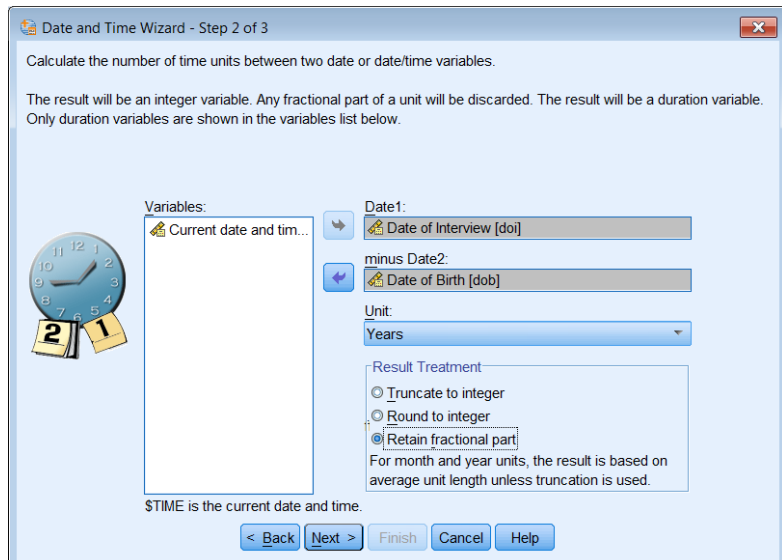


Next choose whether the calculation involves one or two date variables.



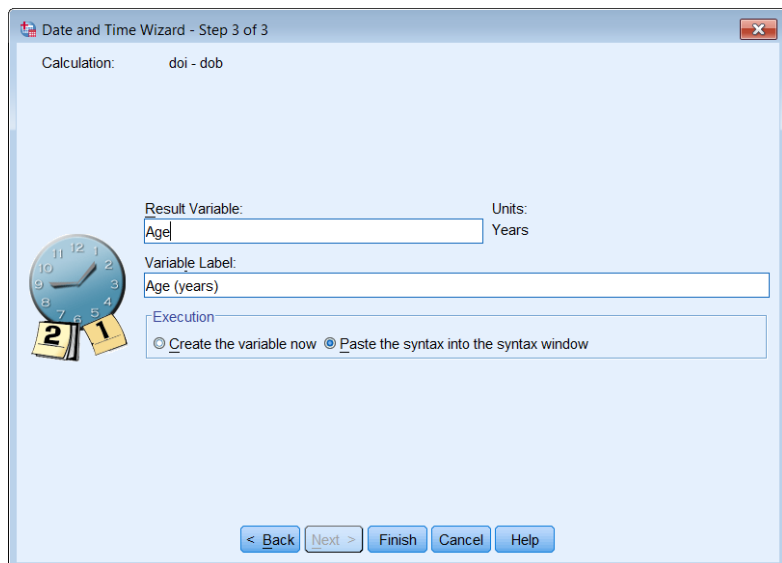
Click **Next**

Select the variables to be used for *Date1* and *Date2* (*Date of Interview doi* and *Date of Birth dob*), the *Unit* required, *years*, in this example and whether an Integer or Decimal result (*Result Treatment*) is required.



Click **Next**

Note SPSS has an in-built *Current Date and Time* variable called *\$TIME* that is available.



The syntax generated by the Date and Time Wizard is shown below. It uses a function called *time.days*.

* Date and Time Wizard: Age.

COMPUTE Age=(doi - dob) / (365.25 * time.days(1)).

VARIABLE LABELS Age "Age (yrs)".

VARIABLE LEVEL Age (**SCALE**).

FORMATS Age (F8.2).

VARIABLE WIDTH Age(8).

EXECUTE.

Counting

SPSS allows you to count occurrences of a response. Suppose you wish to know how many reasons respondents gave for not taking exercise to three questions (*noex1*, *noex2*, *noex3*). Did respondents give no, one, two or three reasons for not taking exercise?

Select **Transform | Count Values within Cases**.

Enter noexcnt in the *Target Variable* text box.

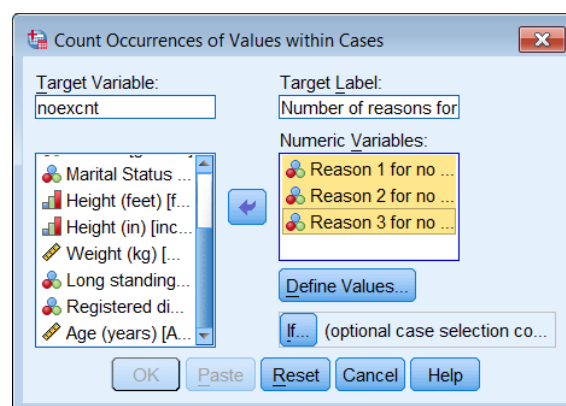
Enter 'Number of reasons for no exercise' in the *Target Label* text box.

Highlight *Reason 1 for no exercise [noex1]*.

Press <Shift> and **Click** on *Reason 3 for no exercise [noex3]*.

(This highlights all three variables sequentially.)

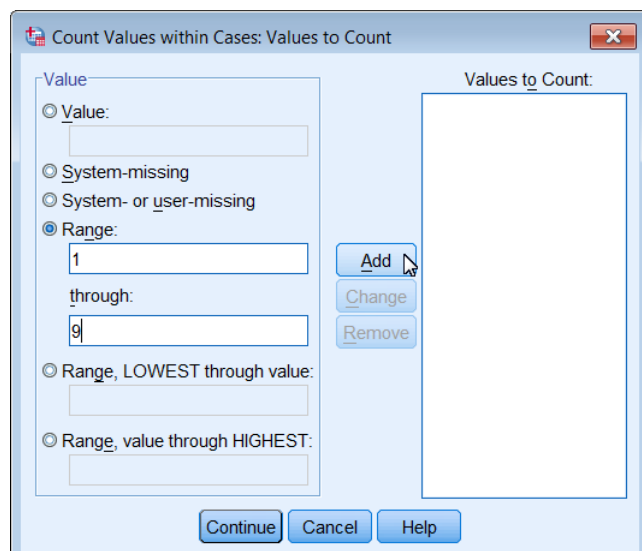
Press  to add these variables to the Variables list box.



The **Define Values...** button is used to enter the values of each variable that are valid, that is the values you want to Count.

For this example a valid response to the three questions (*noex1*, *noex2* and *noex3*) is recorded by values between one and nine.

Click on **Range** in the *Value* panel and enter 1 to 9 and **Add** this to the list as shown.



Count can create a new variable in several ways. It depends on the option selected.

Value	Counts the occurrence of the value given. This is the only option for string values.
System-Missing	Counts the values of the system-missing value.
System or user-missing	Counts any all missing values, system-missing or user define missing.
Range	Counts values in the given range.
Range, LOWEST through value	Counts from the lowest value encountered to given value.
Range, value though HIGHEST	Counts from a given value to the highest value encountered.

Tips and Hints

Count is more flexible than the Dialogue boxes would suggest. Suppose that you wanted to build a variable, called total, from correct answers to various questions (q1, q2, q3 etc.) and that some answers were given more marks than others. Count can be used for this but only by directly typing the command into the Syntax window. For example:

COUNT total = q1 q1 q2 (1).
VARIABLE LABELS total 'Total Marks'.

The Count command would give two marks for answering 1 to question 1 (q1) correctly and one mark for answering 2 correctly, where a value of 1 is a correct answer and any other value is considered to be wrong. The alternative way of doing this, but involving more typing would be:

COMPUTE total = 0.
IF (q1 = 1) total = total + 2.
IF (q2 = 1) total = total + 1.
VARIABLE LABELS total 'Total Marks'.

IF - Combining Logical and Relational operators

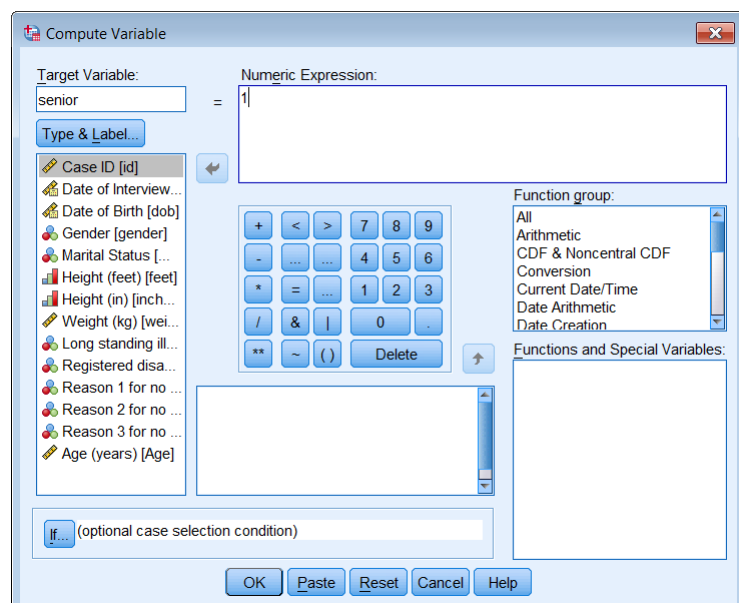
Suppose that you wished to calculate a new variable to identify Senior Citizens in a study. Create a variable *senior* that will equal 1 if the respondent is over 65.

Select **Transform | Compute**.

Click on **Reset** to clear any existing contents of this dialogue box before starting.

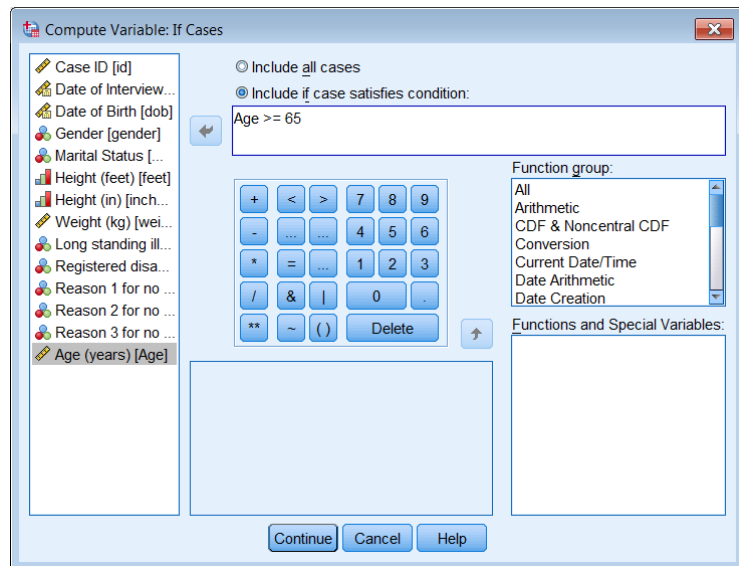
Use the *Compute Variable* dialogue box to create the Target Variable to *senior*, and enter 1 in the Numeric Expression panel.

Click the **If...** button to display the *Compute Variable: If Cases* dialogue box.



The ***Include if case satisfies condition***: radio button must be checked for the logical condition within this window to apply. When ***Include all cases*** is chosen any conditional expression will be ignored.

Build or enter the expression:
age >= 65



The SPSS syntax for this is:

IF (Age >= 65) senior=1.
EXECUTE.

Cases where age is less than 65 will be given the system missing value.

Recode

Suppose that you want to recode Age of Respondents into Age groups. This can be done using the Recode command. The Recode command is comparable to saying , "If the value is now this, make it equal to that".

Select:

Transform | Recode into Different Variables

Click on ***age*** in the list of variables.

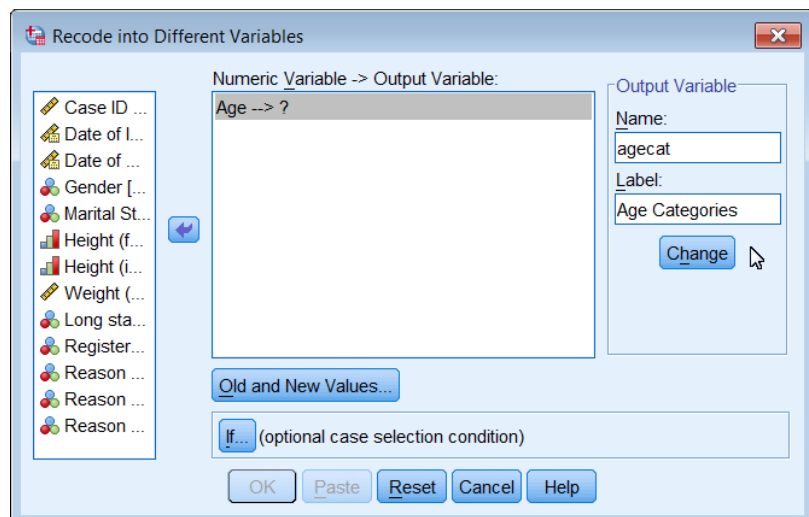
Click on  .

Click on the ***Output Variable- Name*** text box and enter **agecat** .

Click on the ***Label*** text box and enter **Age Categories** .

Click on **Change**.

Click on the **Old and New Values** button.



Always take care of missing values first.

Click on *System- or user-missing* in the *Old Value* panel.

Click on *System-missing* in the *New Value* panel.

Click on the **Add** button.

Click on *Range, LOWEST through value* button (Old Value panel), enter 25 in the text box.

Click on *Value* (New Value panel) enter 1.

Click on **Add**.

Click on *Range*:(Old Value panel), enter 25.01 in the first text box and 65 in the second text box.

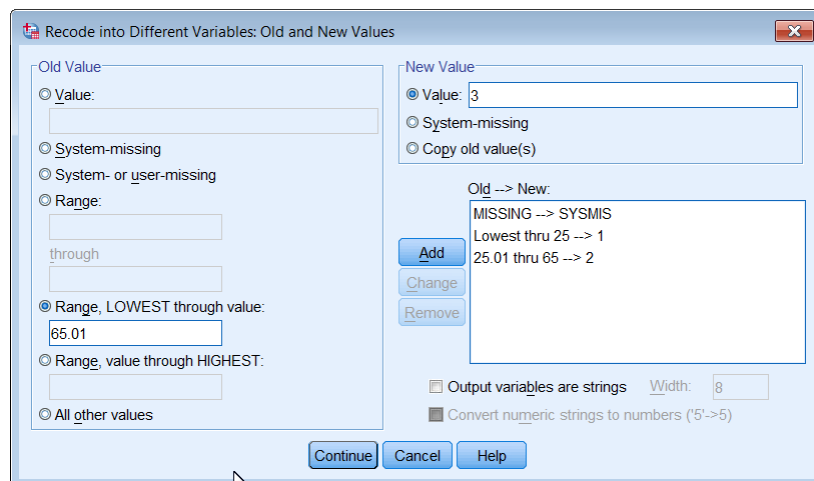
Click on *Value* (New Value panel) enter 2.

Click on **Add**.

Click on *Range, value through HIGHEST* button (Old Value panel), enter 65.01.

Click on *Value* (New Value panel) enter 3.

Click on **Add**.



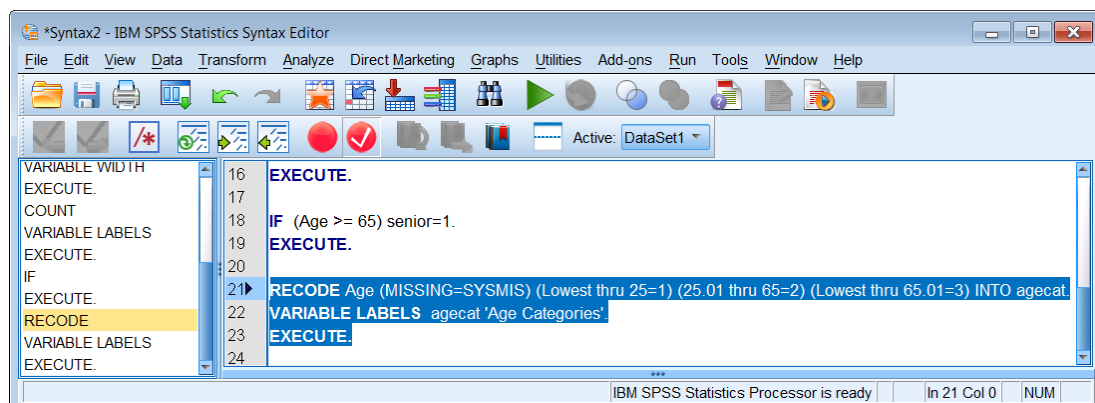
Click on *All other values* button (Old Value panel).

Click on *Copy old value(s)* (New Value panel).


Click on **Add**.

Click on **Continue**.

Click on **Paste**.



Switch to the syntax window and highlight the text:

Click on the Run highlighted command syntax button  .

Switch to the Data Editor window by clicking in it, select the *Variable View* sheet.

Select the *Agecat* Values cell and click on  .

Add suitable *Value Labels* for the *agecat* variable (look at the recode command).

16 Listing Data Values

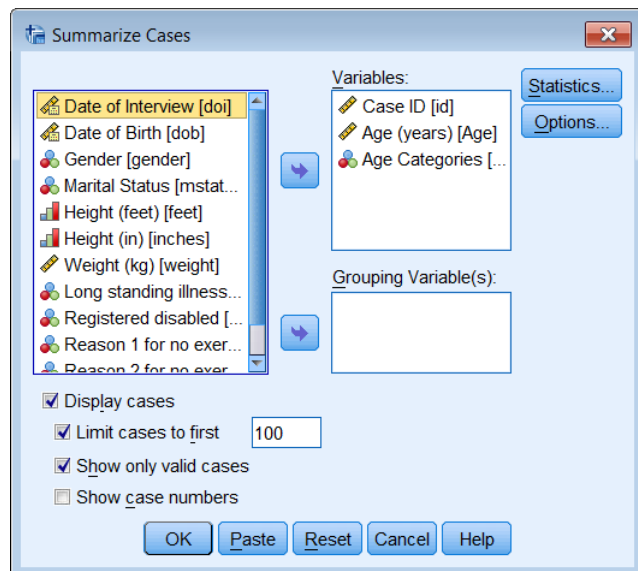
The Summarize command displays the values of selected variables for selected cases or all cases.

Use **Analyze | Reports | Case Summaries** to list data when creating new variables so that you can confirm SPSS has calculated what you expect.

In this example *id*, *age* and *agecat* are summarised to check the recoding as shown in the section of Output below.

Case Summaries^a

	Case ID	Age (years)	Age Categories
1	26	59.05	2.00
2	29	36.15	2.00
3	41	64.04	2.00
4	59	17.47	1.00
5	69	29.37	2.00




17 Saving modified data

After modifying data/variables (by recoding, etc.) and/or creating new variables, you should always save the updated SPSS system file. It is always recommended that you create a new file, for example, Health4.sav. If you save the SPSS data to an existing file (the one you opened initially) and something had gone wrong during your SPSS session, perhaps you accidentally deleted some important variables, you will have to recreate, or re-enter this information.

To save the file, make the SPSS Data Editor the active window (click on it) and then select: **File | Save As** and give the file a new name, e.g. Health4.sav, click the **Paste** button, switch to the syntax window and run the command:

```
SAVE OUTFILE='C:\Users\tnb\Desktop\health4.sav'  
/COMPRESSED.
```

DO NOT Click on the save button  in the Data Editor window as this will try to replace the original data file *soton11.sav*, which you started with.

18 Charts

The SPSS Graphs menu provides three options to create graphs, *Chart Builder*, *Graphboard Template Chooser* and *Legacy Dialogs*. We will be using Chart Builder.

This section will deal with creating charts and plots using the Chart Builder. We will assume that you have the data file *Health4.sav* open.

The Chart Builder allows you to drag elements and variables on to a preview of the type of chart you want to create.


Note: Charts and plots can be created from some statistical procedures on the Analyze menu.

19 Bar charts

Suppose you want to create a bar chart of Marital Status

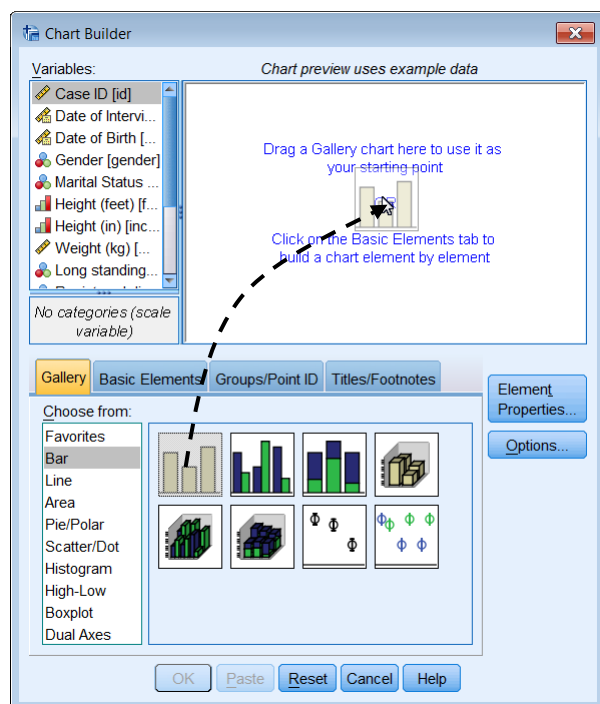
Select **Graphs | Chart Builder**. The dialog box opposite is shown.

Make sure that the *Gallery* tab is showing. Select *Bar* from the *Choose From:* list.

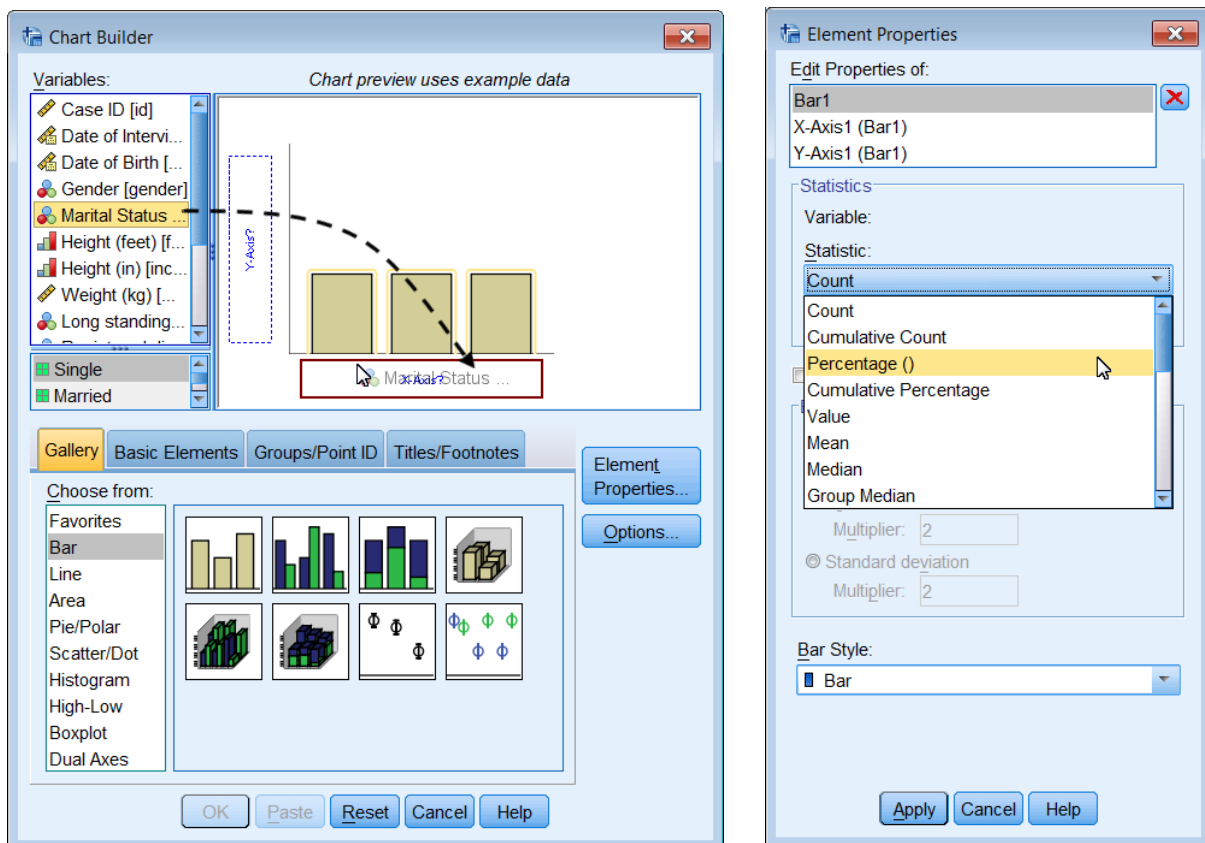
Drag the Simple Bar icon  on to the chart preview area.

A preview of the type of chart is displayed.

An Element Properties dialog will open, see next page.



Drag Marital Status from the Variables list on the X-Axis 'Drop Zone' X-Axis?



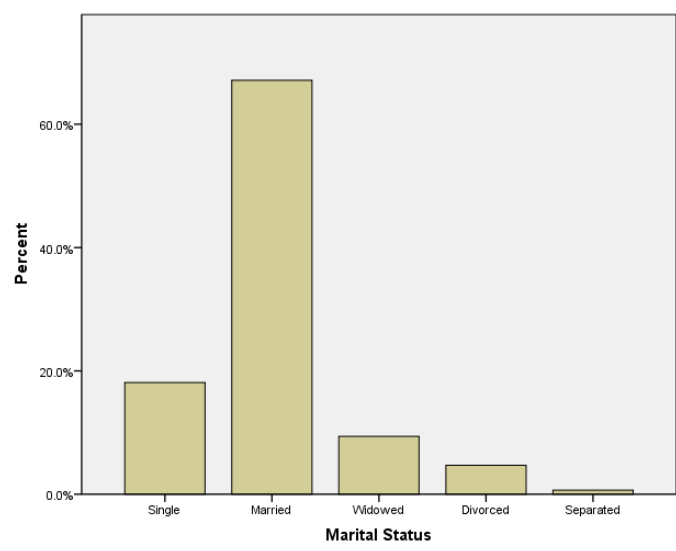
Select the Elements Properties dialog and ensure that in the *Edit Properties of* list **Bar1** is selected. Use the *Statistic* drop down list to select **Percentage()**. Click **Apply** to make this change.

An important point is that the **OK** and **Paste** buttons are not enabled until the Chart Builder has enough information to generate the chart.

Click on Paste, move to the Syntax1 window, highlight the commands (starting from the line *Chart Builder and ending with ENDPGL.) and **Click on** to run them.

The Statistics Viewer window will display the chart opposite.


Charts can be edited by double clicking on them in the Statistics Viewer. They can be exported as graphics files or copied to other applications (Word, PowerPoint). This will be described later.



Clustered bar chart

A clustered bar chart is the diagrammatic equivalent of a cross tabulation. Suppose you want to create a chart of Marital Status by Gender.

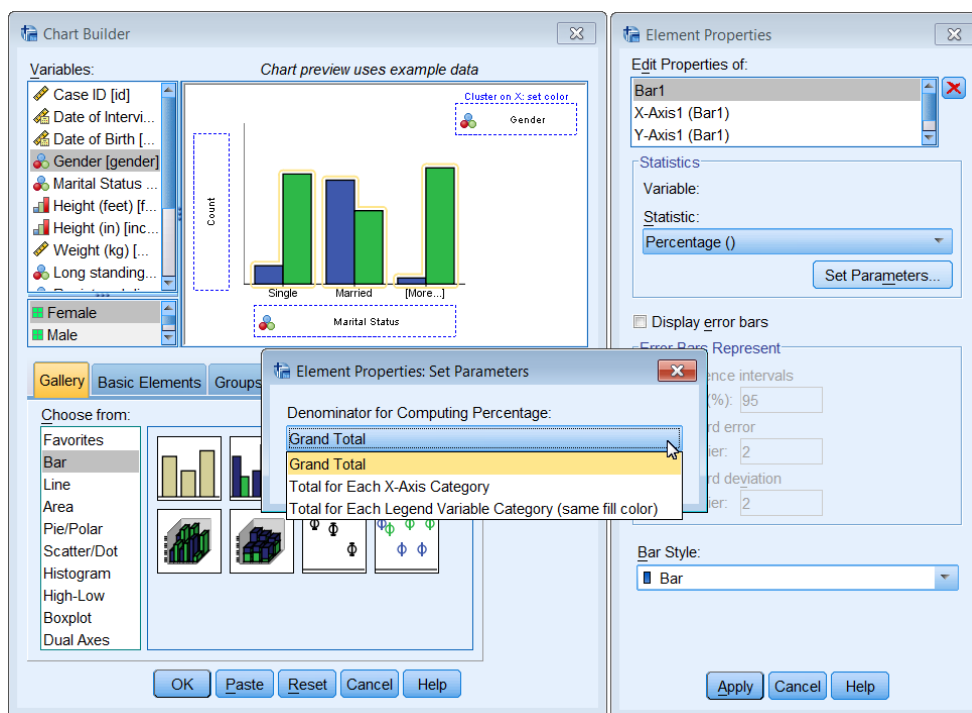
Select **Graphs | Chart Builder** the dialog will show the previous chart you created.

Drag the Clustered Bar icon  into the Chart preview. There is no need to Reset. Drag **Gender** from the Variables list to the **Cluster on X set color** 'Drop Zone'.

To use Percentages instead of Counts, change the **Statistic** drop down to **Percentage ()** in the Element Properties window.

Click the **Set Parameters** button to set the **Denominator for Computing Percentage** to **Grand Total**. There are various parameters which you chose depending on how you want to represent the data. Click to the Help button for an explanation of these.

Click **Continue** and then **Apply**.



If you want to add a title to the chart, select the **Titles/Footnotes** tab.

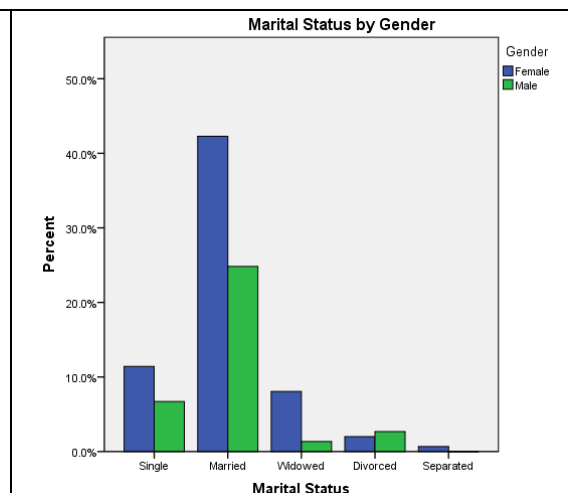
Tick the **Title 1** box.

Move to the **Elements Properties** dialog, select the **Content** panel and enter your title '**Marital Status by Gender**'

Click **Apply**

Click on **Paste** and **Run** the command from the syntax window.

The chart opposite will be produced.



Note: You may prefer to omit the title and add it as text in your Word or PowerPoint document above the imported image.

20 Scatterplots - XY graphs

To create a Scatterplot chart select **Graphs | Chart Builder**.

Press the **Reset** button to clear any previous charts.

Make sure that the **Gallery** tab is showing.

Select **Scatter/Dot** from the **Choose from:** list.




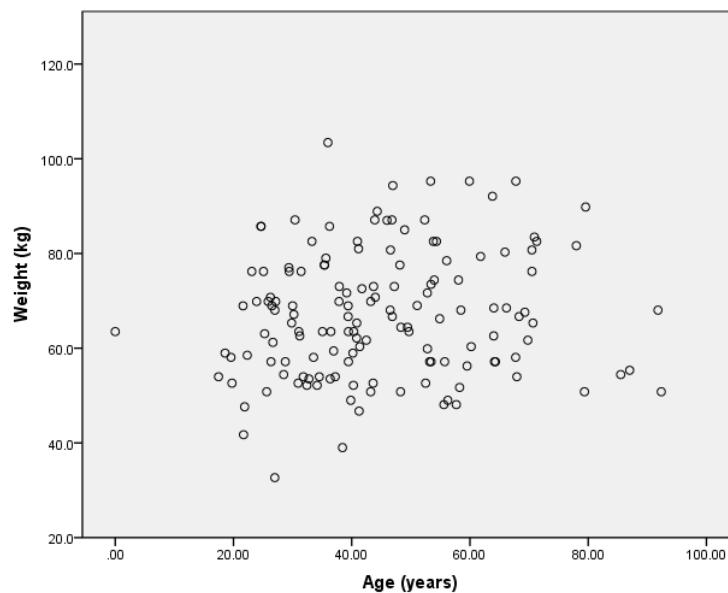
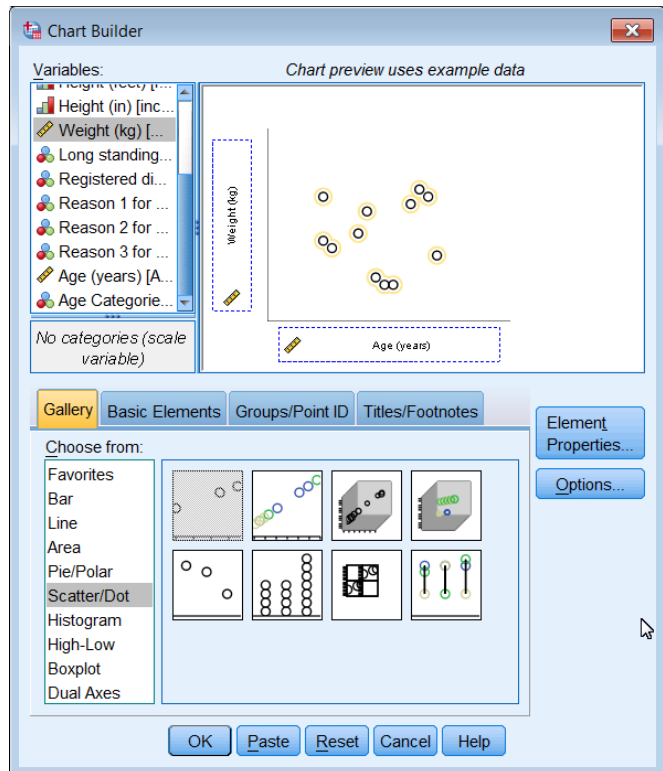
Drag the Simple Scatter icon onto the chart preview area.

A preview of the type of chart is displayed.


Drag **Age** onto the X Axis Drop zone


Drag **Weight** onto the Y Axis Drop zone

Either **Click Paste** and then  from the syntax window, or **Click OK**.



Note: You can reduce or remove the decimal places on the axes scales before generating a chart using the Data Editor. In Variable View sheet and change the Decimals for each variable to zero

If you wish to identify males and female in the scatter plot, **Click** on the **Dialog Recall** button  on the toolbar and select **Chart Builder** from the list.

Choose the Gallery tab and drag the Grouped Scatter icon  on to the Chart Preview area. Drag **Gender** from the Variables list to the **Set color** 'Drop Zone'.

Paste and **Run** this command.

This will plot female and male data points as blue and green open circles.

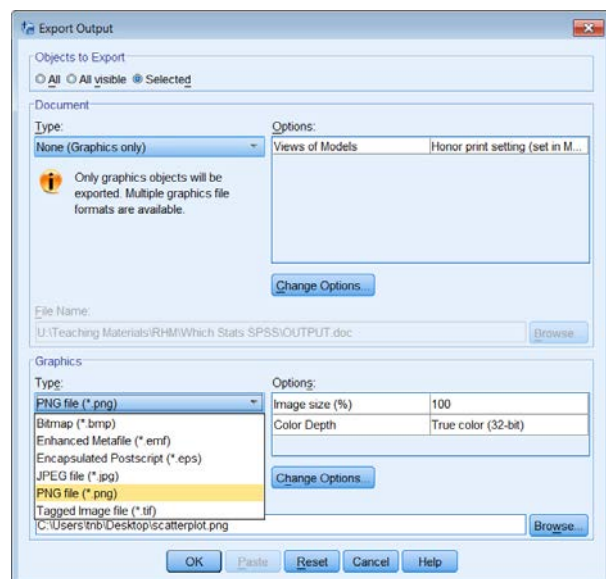
21 Incorporating SPSS charts into Word or PowerPoint

Note Make all the modifications/changes to your chart before transferring it to another application. Although Word and PowerPoint have some editing capability, they are less flexible!

There are two options for placing a chart in a document.

1. **Copying.** From the SPSS Statistics Viewer select the chart, select **Edit** | **Copy**. Move to the other application (such as Word or PowerPoint) and select **Edit** | **Paste**.
2. **Export** the chart, by selecting it in the SPSS Statistics Viewer. Select **File** | **Export** from the menu.
 - i. Choose which **Objects to Export**:
 - ii. From the **Document Type** drop down select **None (Graphics Only)**
 - iii. From the Graphics **Type** select **PNG File (*.PNG)**. Better than JPEG
 - iv. **Click** on **Browse** in the **Root File Name** panel to locate the folder and specify a name for the export file in the text box

Click OK



Saving a file as an image removes your ability to edit the chart again using SPSS. If you think that you might wish to edit the chart again, save the contents of the SPSS Statistics Viewer which will include the chart. The SPSS Statistics Viewer output document can be reopened and any Chart opened by the Chart Editor and changes made. However SPSS Statistics Viewer files (*.SPV) can be very large, so it is best delete any unwanted charts before saving the SPSS Statistics Viewer content.

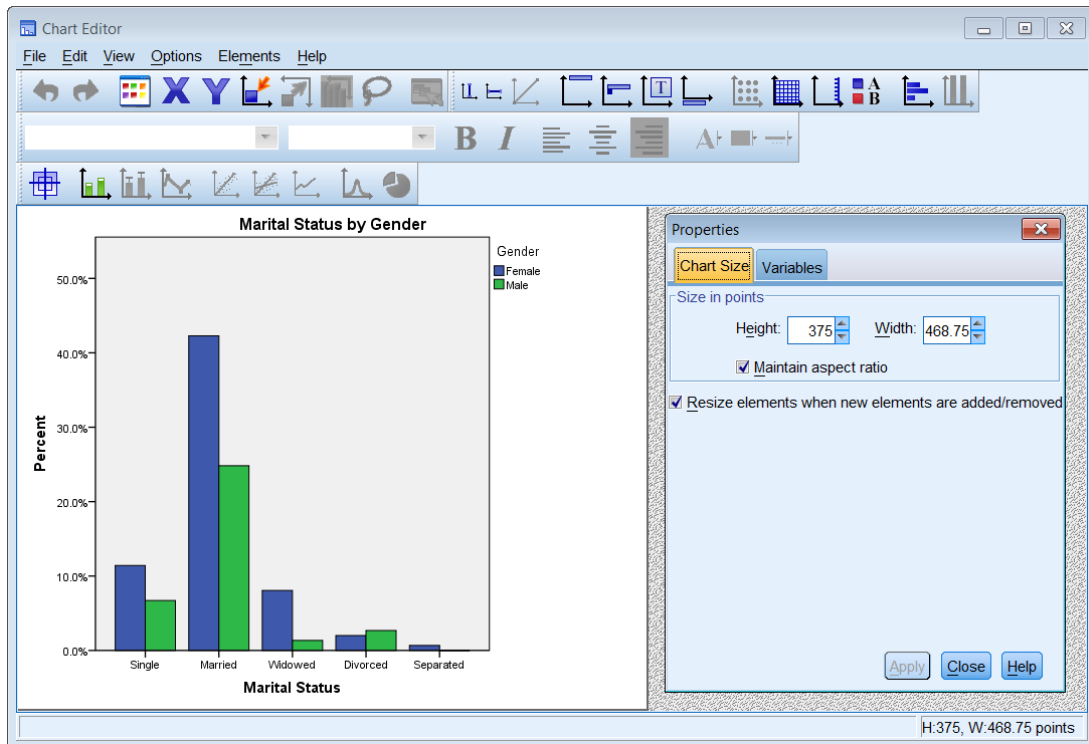
22 Saving Charts

It is recommended that you save charts as individual files if you intend to use them in other applications (Word or PowerPoint).

23 Modifying Charts

Charts may be modified by **double clicking** the chart in the SPSS Statistics Viewer. This opens the Chart Editor window, revealing the chart menu and buttons that may be used to modify the chart.

The figure below shows a Clustered Bar chart created previously for Marital Status by Gender displayed in the Chart Editor window. A properties dialog is also shown. The Properties dialogue displays different content according to which is the currently selected object in the chart.



The options for each menu item are described on the next page. They provide general and chart specific functions.

Chart Editor Menu options

File You can Save Chart Template (save a chart to use as a template on other charts), Apply Chart Template and Export Chart XML (to be read by or incorporated into other applications).

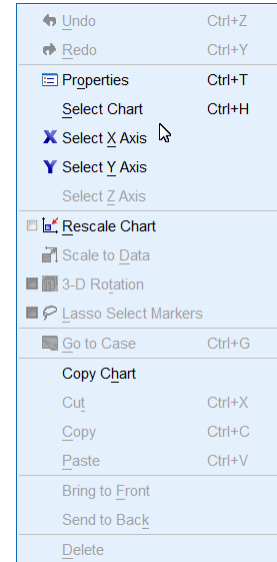
Edit Provides the standard **Undo** and **Redo** options and the Cut /Copy and Paste functions for some chart elements.

Properties displays the properties dialogue box which is used to change the properties of the currently selected element in the Chart Editor. The properties dialogue box can also be opened by

clicking  on the Chart editor toolbar.

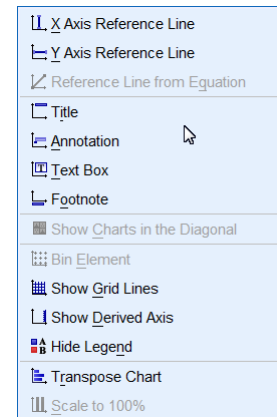
The tabs displayed vary according to the chart object that is currently selected.

Copy Chart copies the chart to the Clipboard from where it can be pasted into another application.

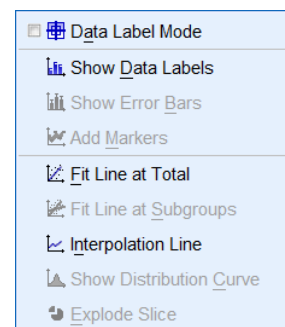


View From the View menu, you can control which toolbars are displayed and the size of the buttons on them.

Options Provides features to add text and lines to the chart. Many of these functions are provide on the Toolbar as well.



Elements Used to show Data Labels, Error Bars, Line Markers or add regression lines to a chart.



Help Provides the same access to Help as it does throughout the system.

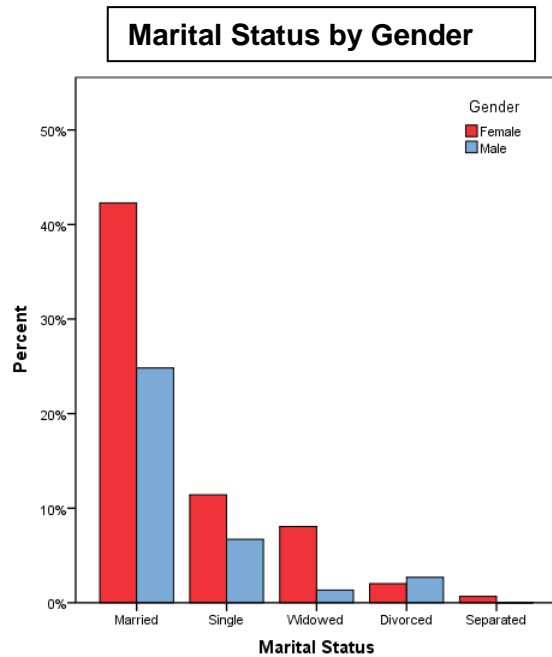
Modifying Chart Elements

The figure opposite shows the Clustered Bar chart created previously for Marital Status by Gender (see page 52) that has been modified using the Chart Editor.

A chart is made up of a number of Elements or Objects. Each element can be edited by clicking on it to select it.

The selected element is identified by a yellow border around it.


Note The chart editor can be slow to respond when you click on an object to select it. Be patient; count to five if it is slow!



To modify any chart objects either:

Double-click on the object in a chart, and the Properties window will appear;

or

Single right-click on the object and click the Show Properties Window icon, ;

or

Single-click on text (such as a label) and click again on the text to change the text content

or

Use select **Edit | Properties** on the Chart toolbar.

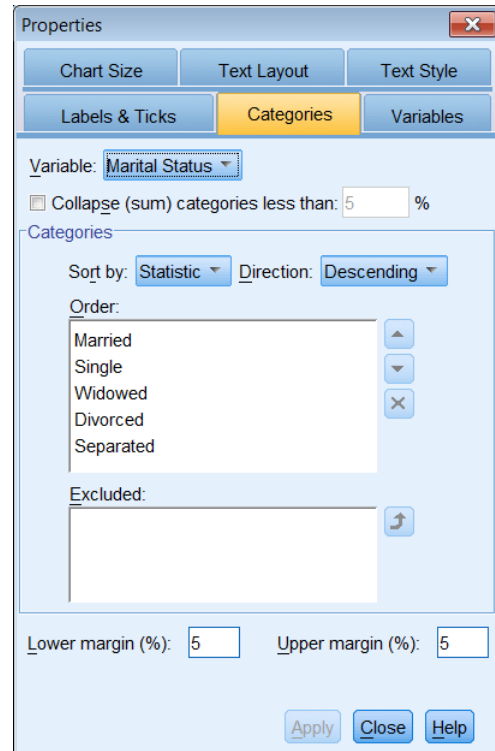
For example to change the order in which the categories for marital status are displayed, **double click** on any bar.

Select the **Categories** tab in the Properties window.

Select the **Variable** of interest, **Marital Status**








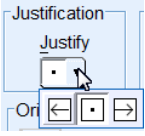
Select your **Sort by**: option, **Statistics**

Select the **Direction**: for the sort, **Descending**



Note: Whenever you make a change to an object in a chart you must click **Apply** for it to take effect.

The Marital Status by Gender clustered bar chart (shown above) was modified as follows:

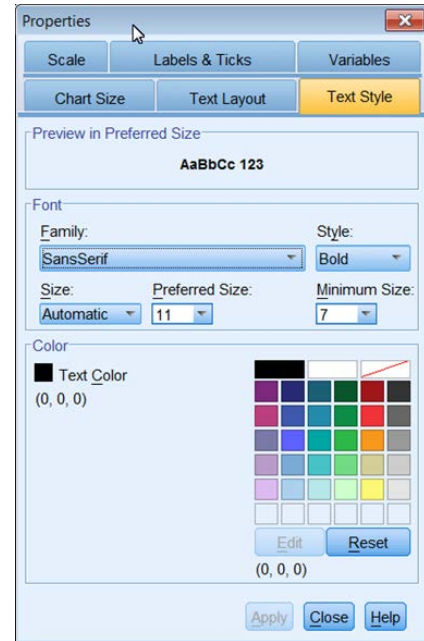
Chart window	Double-click on the chart in the SPSS Statistics Viewer to open the Chart Editor window.
Chart Title	<p>You are advised to add a chart title once the image has been imported into Word or PowerPoint.</p> <p>To remove a title, click on it, a yellow box with handle will appear. Press the Delete key.</p> <p>To add a title using the Chart Editor, click the <i>Insert a Title</i> icon  or select Options Title from the menu. Overtyping the word Title with <u>Marital Status by Gender</u>. (Make two single clicks on Title if it is not selected) Select the Properties dialog, Use the <i>Text Style</i> tab to change Font or Size. Click Apply</p>
Removing a Missing Data bar	<p>If a bar for missing data is displayed, it may be removed by highlighting the bars on the chart (click on any bar). The <i>Properties</i> window should be displayed, if the window does not appear click the Properties, , icon.</p> <p>Select the <i>Categories</i> tab, use the <i>Variable</i> drop down  to select the variable with the missing item (e.g. Marital Status).</p> <p>Select the “Missing” item from the items in the <i>Order:</i> list. It may be the first item with no text. Click, , to move it to the <i>Excluded:</i> list, Click Apply.</p>
Reordering the bars	<p>This was described on the previous page. To order bars according to your own requirements, use the <i>Properties</i> dialog and <i>Categories</i> section, change <i>Sort by:</i> to <i>Custom</i>, reorder the categories (to married, single, etc) using the up and down arrows .</p> <p>Click Apply.</p>
Moving the Legend	<p>It can be difficult to select the legend. Click to select it. The legend is selected when a yellow rectangle surrounds it. If only a part of it is selected, try again.</p>  <p>Position the mouse over one of the yellow lines and drag the legend to its new location.</p>
Changing axis titles	<p>To edit the title of the X or Y axis, single-click on the axis title to select it. Click again on the text to change the text of the Axis title.</p> <p>To change the title justification to Right Aligned, click on  to display the Properties dialog (if not already shown), select the <i>Text Layout</i> tab and select the right arrow in the Justification panel. Click on Apply.</p> 

Changing axis styles

Use the **Text Style** tab (opposite) to change the Font and Size of text.

To change the scale, move the cursor over one of the scale numbers and click.

Use the **Scale** tab in the **Properties** dialog (not shown) to change the numbers and increments on an axis. Untick the options and replace them with your own values. Click the **Apply** button.




Changing bar colours


Clicking on a bar / category to select different items. **The number of clicks is important.**

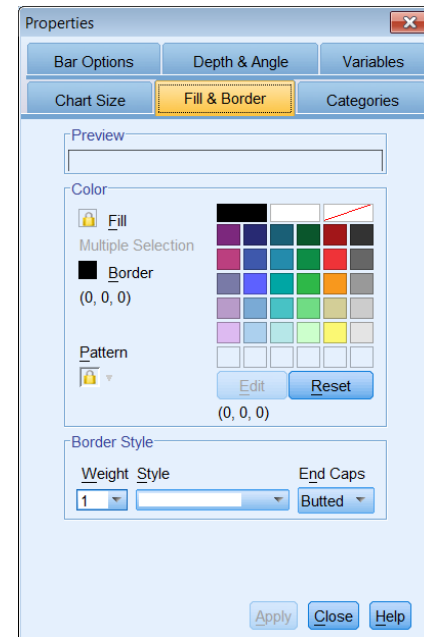
One click selects all bars in the chart – they show a yellow outline
A second click (not double click) selects one of the categories.
A third clicks select a single bar/category.

When a single category has been selected, clicking on another category selects that category.
To select all bars/categories again, click on another chart element and then click back on a bar/category.

Bar colours can be changed by selecting the **Fill & Border** tab in the Properties window. Using two clicks to select a bar, selecting the Fill colour box , and choosing a new colour from the palette. Click **Apply**.

In the example blue was changed to red, a green bar was clicked (to select all green bars) and these were changed to blue,


If the colour or border box displays  then you may have selected multiple objects (more than one bar); colours can only be changed for single objects.



If you wish to display the data as counts rather than percentages then you must specify this when creating the chart from the **Graphs | Bar...** dialogue box. Counts is the default option for the Y axis of a bar chart but percentages may be more helpful when comparing groups.

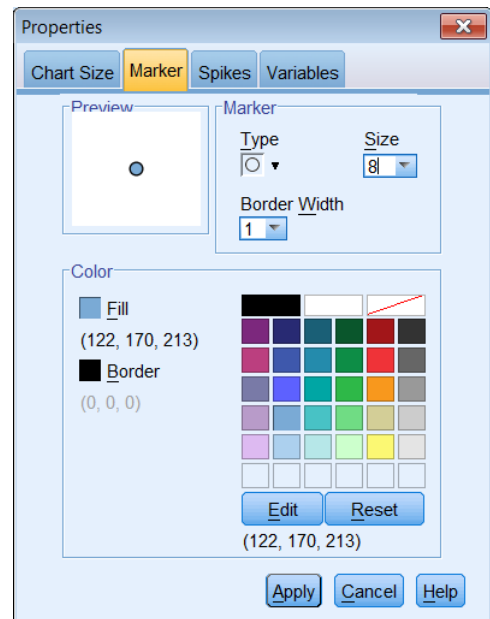
Modifying a scatter plot

Open the Chart Editor window by **double clicking** on the chart (shown on page **Error! Bookmark not defined.**).


To change the symbols used to plot each case, **double-click** on any of the plotted points (open circle) and the *Properties* window will appear. To change the symbol to a closed circle, click on the **Fill** box , then select a colour from the palette. You might wish to change the size of the symbol as well.

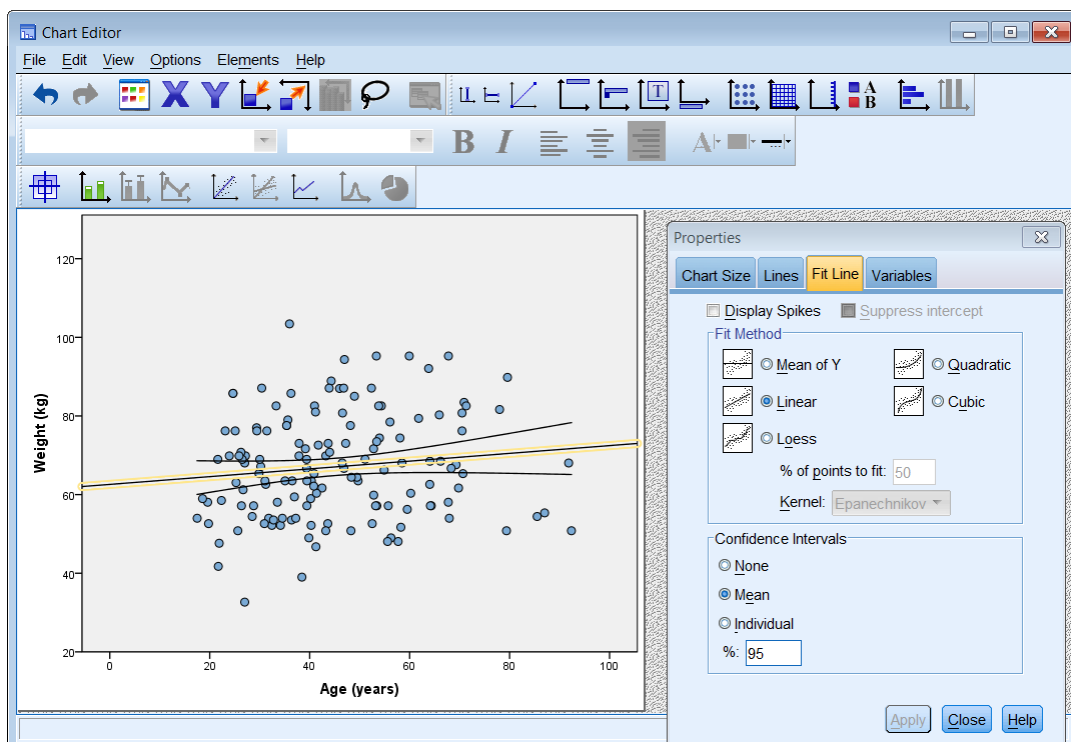
Click on **Apply** to make the changes.

Note: If you wish to change the fill colour or shape (Type) identified by a grouping / marker variable, e.g. Gender, select the marker for that group from the legend window, or click on a single data point twice so that only points of that colour are selected. Then change the properties of that group and click Apply.



Adding a regression line

To add a regression line through the points, click on the *Add Fit Line at Total* button.  The default Fit Method is to fit a linear regression line. To add confidence intervals for the line, use the *Properties* window, **Fit Line** tab to select the **Mean** in the Confidence Intervals panel (use 95% CI) and click the **Apply** button. Or, select the **Individual** option (to add the prediction limits) and click the **Apply** button. The figure below shows the regression line with the 95% confidence interval for the regression line. The legend has been deleted and decimal digits removed from the axes scales in the chart below.




*** The next section of the notes describes how to use various statistical tests in SPSS. You may also find the web page **Which Statistical Test?** useful.
www.som.soton.ac.uk/learn/resmethods/statisticalnotes/which_test.htm

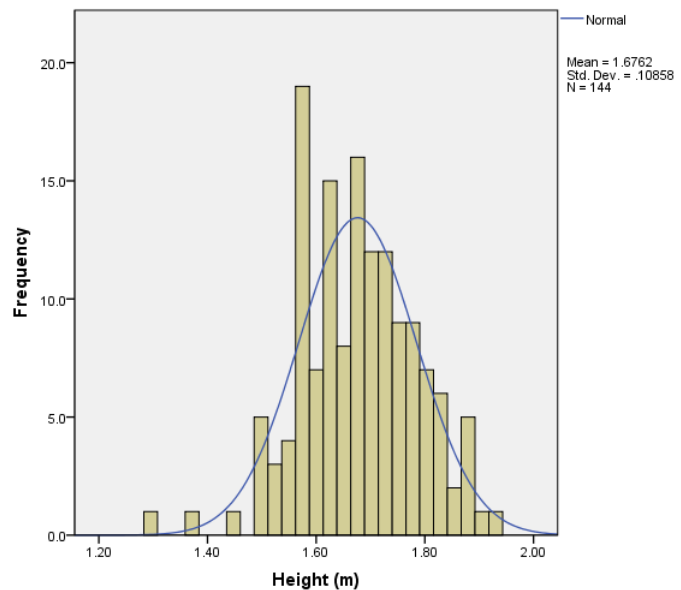
24 Comparing Means and Medians

Suppose you wish to compare the means heights of females and males in our study. If you want to use parametric statistical procedures, then your assumption is that the data you have sampled was taken from a Normal distribution.

You can check your assumption by producing a Histogram with a Normal curve superimposed (**Graphs | Chart Builder** -select Histogram from the


Chooses from panel and drag  into the Chart preview area. An Elements properties dialogue will be displayed.

Drag **Height (m)** from the variable list to the X-axis in the Chart preview window. Tick the **Display normal curve** box in the **Elements Properties** dialogue and click **Apply**.

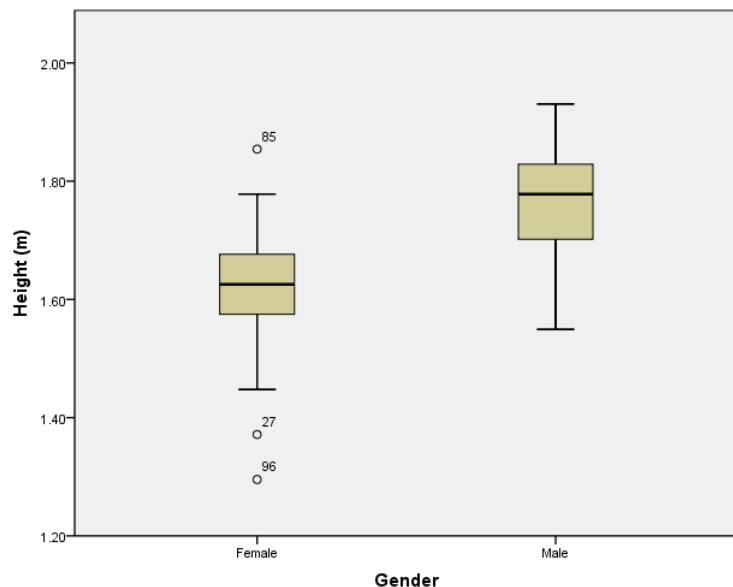


Before carrying out any statistical calculations, you should plot the data so that you can visually compare the data for the groups for any differences.

One approach could be to plot separate histograms for the groups (female and male). A better way to view the data is with a Box and

Whisker plot .

In the diagram opposite the Medians, represented by solid horizontal band, will be similar to the mean values for Normally distributed data. In this chart the medians are different and the boxes representing 50% of the data for each group and do not overlap. So you can hypothesise from this that the mean heights of females and males are different.





Two Independent Samples: t-test

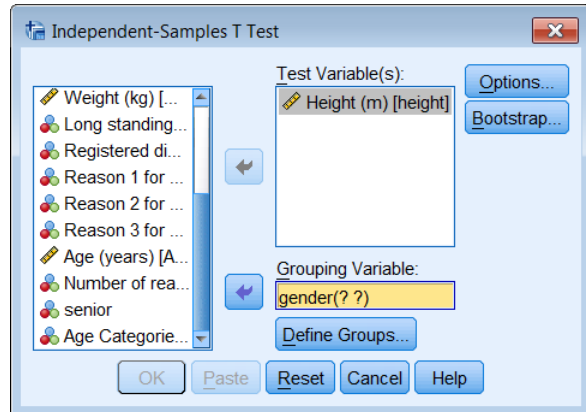
To test whether the observed difference between the mean heights of the female and male subjects in our study could have arisen by chance, select:

Analyze | Compare Means | Independent-Samples T Test

An Independent-Samples T Test dialogue box is opened.

Click on *Height (m) height* and the upper  to add this variable to the *Test Variable(s)* list.

Click on *gender*, the variable that defines the two categories and on the lower  on to add it as the Grouping Variable.



Gender is displayed as *gender(? ?)* in the *Grouping Variable* box.

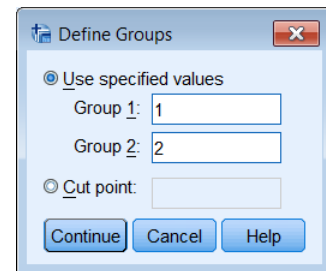
Click on the **Define Groups** button.

The Define Groups dialogue box allows the values that specify the two groups to be entered.

Enter 1 in the Group 1 box, press <tab>

Enter 2 in the Group 2 box.

Click on **Continue**.



The Grouping Variable box of the Independent-Samples T Test will display *gender (1 2)*.

The **Options** button in the Independent-Samples T Test dialogue box displays an *Options* dialogue box that allows the value for confidence intervals to be altered (default = 95%), value labels to be suppressed on the output and controls how missing values should be treated (List or Pairwise). These do not need to be altered.

Paste and **Run** the command.

Output from SPSS t-test is shown below. Two tables are produced, one for Group Statistics and the other for the Independent Samples Test.

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Height (m)	Female	93	1.6261	.08751	.00907
	Male	51	1.7675	.08035	.01125

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower		Upper
Height (m)	Equal variances assumed	.069	.793	-9.541	142	.000	-.14139	.01482	-.17069	-.11210
	Equal variances not assumed			-9.782	110.735	.000	-.14139	.01445	-.17004	-.11275

Before viewing the test statistics you should have decided if the variances (standard deviations squared) of the two groups are similar (equal) or not. If the variances are similar, the top row of the table (Equal variances assumed) is used. If the variance are not similar the bottom row (Equal variances not assumed) is used. The important columns are the Mean Difference (-.14139) and the 95% Confidence Interval of the Difference (-.17069 to -.11210) together with the Sig (2-tailed) column which displays the P value (.000).

Note the negative values for the Difference and the Confidence Intervals of the difference occur because the female group, which is lighter, is the first group. If the male group had been the first group the values would have been positive.

Comparing Two Medians: Mann-Whitney test

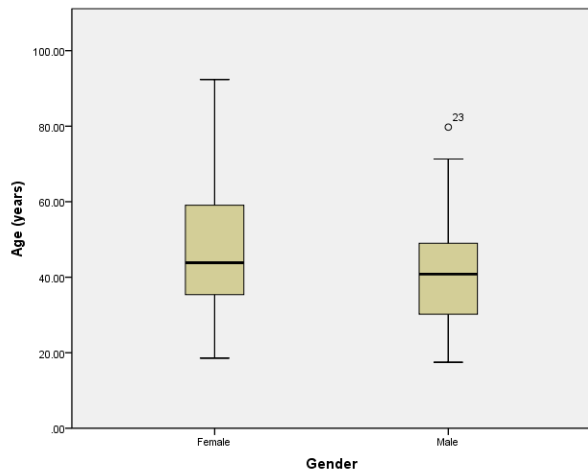
The Independent-Samples T Test used above assumes that the data are from populations that have a 'normal distribution', or that the sample sizes are large enough so that the distribution of the sample means is 'normal'. When the assumption of normality cannot be used, alternative techniques must be employed. These techniques are known as 'distribution-free' or 'non-parametric' statistics. They rely on the order or rank of the data rather than its distribution, 1 is smaller than 2; 4 is larger than 3. The fact that their differences are the same magnitude (2-1= 1 and 4-3 =1) is not a consideration.

In our data set you can consider that Age is not from a normal distribution because no case is under 17 years old (the subjects were sampled from a voting list). The distribution is truncated to the left.

Suppose you ask, is there a difference in the Ages of the females and males in this study?

The Box and Whisker plot compares the female and male data. This shows that the medians are very similar and there is complete overlap of the distributions.

Note: From this figure you might conclude that there is no point in testing for a difference because the median are similar. However we will demonstrate use of the procedure.



You can test for a difference between the ages of females and males by using the non-parametric equivalent of the T Test, the Mann-Whitney test. Select:

Analyze | Nonparametric Tests | Legacy Dialogs | 2 Independent Samples

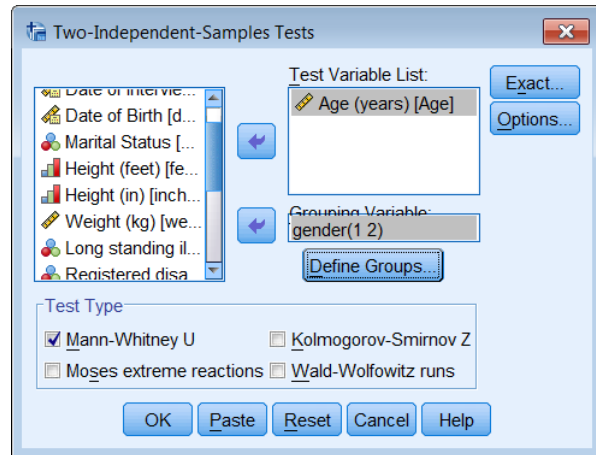
Highlight *Age of Respondent* and click the upper button to add it to the Test Variable list.

Select *Gender* as the grouping variable. Define Groups as 1 and 2.

Paste and Run the command.

	Age (years)
Mann-Whitney U	2008.000
Wilcoxon W	3386.000
Z	-1.782
Asymp. Sig. (2-tailed)	.075

a. Grouping Variable: Gender



Tests statistics given by the non-parametric Two-Independent Samples Test shown above, provide very little useful information. The item of interest in this table is the *Asymp. Sig (2-tailed)*, or P value (.075). This indicates that there is no significant difference between the ages of females and males which could have been predicted by looking at the box plot.

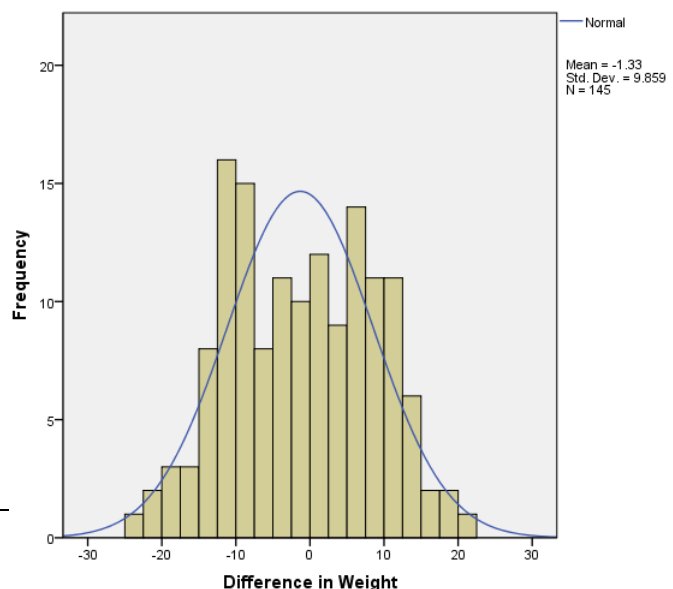
Note: The output from all of the Non-parametric procedures in SPSS is poor. Typically, the only useful statistics the Asymp. Sig (2-tailed). You must use other SPSS procedures to obtain the medians of each group (e.g. Explore); confidence intervals are not calculated but can be obtained using the Confidence Interval Analysis Program, CIA.

Paired samples t-test

Note The following is for illustration only. The data set you are using does not include any paired data.



A paired test is used when the data are matched or paired in some way. Often the subjects are used as their own control For example measuring something 'before' and 'after' an event. The paired t-test calculates the difference between paired data items and tests whether the average difference between paired measurements is different from zero.

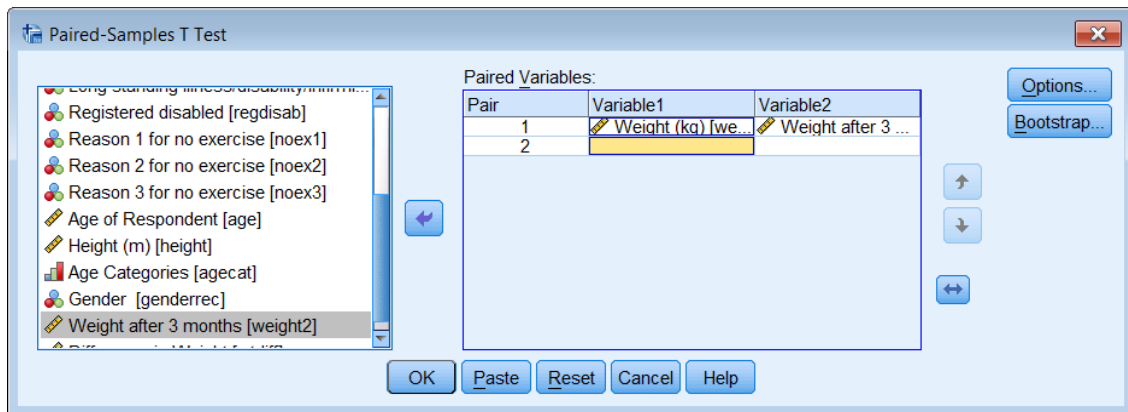
Suppose you measure the weight of subjects, prescribe a new diet and measure their weight three months later. If the new diet reduces weight, the average change in weight of our subjects should be different from zero. This can be visualised by calculating the difference between weights (**Transform | Compute**) and plotting a histogram of the difference (weight - weight2) (**Graphs | Chart Builder - Histogram**) as shown opposite.




Note: The paired Sample *t*-test can be used to test the difference between two variables where the original data are not from a Normal distribution if the distribution of the differences between the two readings approximate to a Normal distribution.

The SPSS procedure would be: **Analyze | Compare Means | Paired Samples T Test**

The data for this procedure must be entered as two variables. These are selected by scrolling the list of variables and **clicking** on a variable and then clicking on  to add it to the **Paired Variables** panel. If you select an incorrect variable, **clicking** on it in the Paired Variables panel (it will be highlighted in blue) and click  to remove it.



The order of the variables (Weight 1 – Weight 2 or Weight 2 – Weight 1) can be set by clicking the  button.

Output for the Paired Samples procedure is shown below. The mean difference of -1.3257 before minus after values) suggests that our subjects lost weight! However, this difference is not significant ($p = 0.108$) and the change is due to random chance. You can tell this because the 95% confidence interval (-2.9940 to .2927) includes zero and also because the Sig (2-tailed) is 0.108.

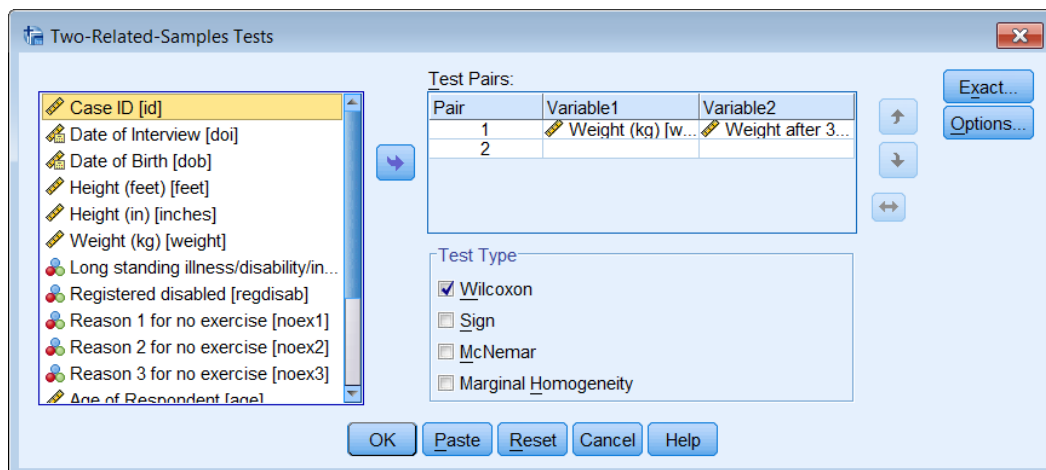
		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Weight (kg) - Weight after 3 months	-1.3257	9.8594	.8188	-2.9440	.2927	-1.619	144	.108

Non parametric Paired Samples: Wilcoxon

The nonparametric equivalent of the paired *t*-test is the Wilcoxon signed-rank test. This considers whether the difference is above or below zero. The procedure would be:

Analyze | Nonparametric Tests | Legacy Dialogs | 2 Related samples

Variables are selected in the same way as for the paired samples t-test procedure.



The only useful information displayed is the Asymp. Sig (2-tailed) (P value), .108.

The calculation is Variable2 – Variable 1 which the reverse of the paired t-test!

Test Statistics^a

	Weight after 3 months - Weight (kg)
Z	-1.606 ^a
Asymp. Sig. (2-tailed)	.108

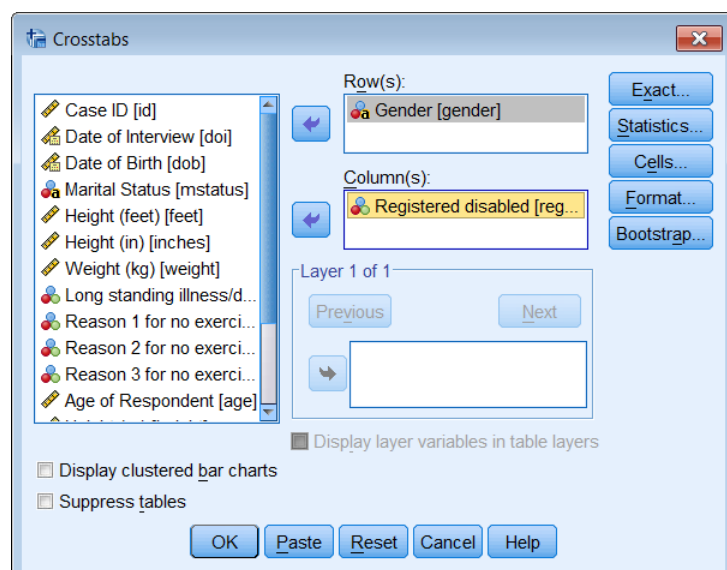
a. Based on negative ranks.
b. Wilcoxon Signed Ranks Test

25 Comparing Proportions 2x2 tables

We are often interested in examining the relationship between two categorical variables. This may be achieved by cross-tabulating them in a contingency table. A useful convention (when appropriate) is for the rows of the table to correspond to the exposure values and the columns to the outcomes. You might ask whether gender is associated with registered disability, so gender is the “exposure” and regdisab (registered disabled) is the “outcome”.

To create the table in SPSS select: **Analyze | Descriptive Statistics | Crosstabs**

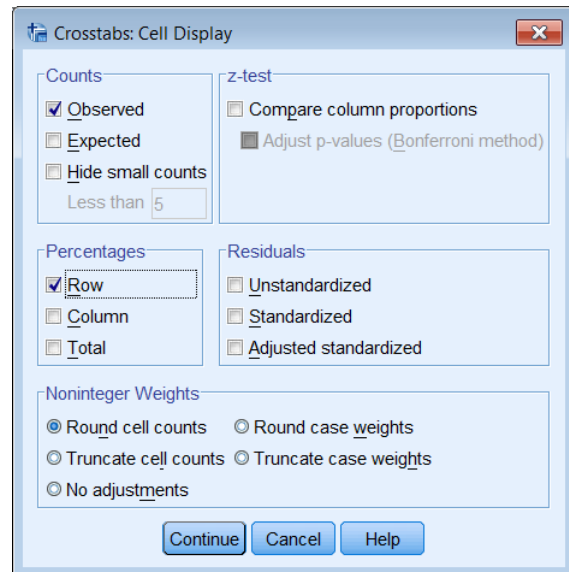
Use *Gender* as the Row variable and *Registered Disabled [regdisab]* as the Column variable.



To improve the presentation and allow more detailed examination of the association by including marginal totals and percentages, click the **Cells** button and check the **Row Percentages**.

Click **Continue**

Click **Paste** and Run the command.




The output is shows that there are 15.8% of females who are registered disabled and 0% of males.

A word of caution the number of Registered Disabled is very small so care must be taken with any statistical testing.

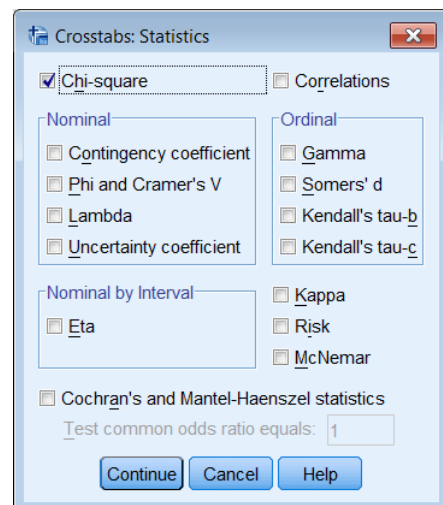
Gender * Registered disabled Crosstabulation

		Registered disabled		Total	
		Yes	No		
Gender	Female	Count	3	16	19
		% within Gender	15.8%	84.2%	100.0%
	Male	Count	0	12	12
		% within Gender	.0%	100.0%	100.0%
Total		Count	3	28	31
		% within Gender	9.7%	90.3%	100.0%

Chi-squared test for a 2x2 table

Suppose you wish to assess the strength of the association between gender and registered disability. Use the Recall dialog button  to recall the Crosstab procedure. Click the **Statistics** button.

Check the **Chi-square** box.
Click **Continue**
Paste and Run



The viewer will display as output:

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	2.098 ^a	1	.148		
Continuity Correction ^b	.680	1	.410		
Likelihood Ratio	3.138	1	.076		
Fisher's Exact Test				.265	.216
N of Valid Cases	31				

a. 2 cells (64.5%) have expected count less than 5. The minimum expected count is 1.16.
b. Computed only for a 2x2 table

Interpreting the output

In this example you have small numbers, SPSS warns you in footnote b of the table above. Therefore the Pearson chi-square statistic which is the χ^2 test statistic is not appropriate in

this circumstance; Fishers Exact test should be used here (see below).
If you had sufficient data you would use the (Asymp. Sig. 2-sided) = 0.148 which is the P-value and observe that is not significant.

Yates' continuity correction

Some statisticians argue that the chi-squared test for a 2x2 table can be improved by using a continuity correction, (Yates' continuity correction). In the output above the *P*-value is increased to .410.

Fisher's exact test

In this example an Exact test should be used to compare two proportions when the numbers in the 2x2 table are very small, leading to an expected value (in any cell) of less than 5. SPSS will report if any cells have expected values of less than 5 in which case the Fisher's exact test *P*-value should be reported. So you should report $P=0.265$.

Other statistics from Crosstabs

Paired Data (McNemar's test)

This is used when you have paired data for before and after studies, for example the presence of a condition (e.g. skin rash) is observed in subjects before and after an intervention. The χ^2 test is not appropriate because the data are not independent.

An example of this is: Seventy seven patients with diabetic nephropathy had one eye chosen at random to be treated while the other eye remained untreated. After five years follow-up the response variable (blind or not) was recorded for both eyes.

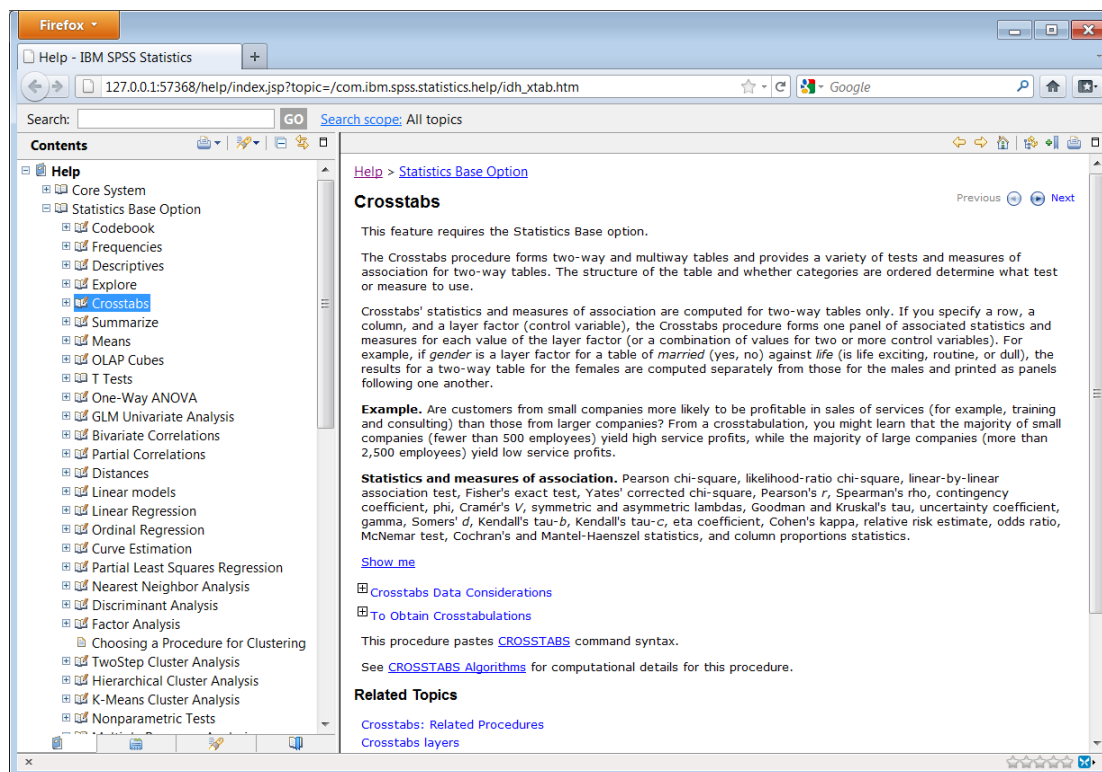
The kappa (κ) statistic

This allows you to assess the extent of the *reproducibility* of a measurement (also known as *reliability*), including:

- Agreement between different observers (inter-rater agreement) using the same measurement technique.
- Agreement between replicate measurements taken at different points in time.

26 Help with Statistical Procedures

To find out more about SPSS statistical procedures, Click the **Help** button in the procedure dialogue window, for example Crosstabs). A short description of the procedure is provided with an example and the statistics than can be calculated.



27 Further Reading

For a greater understanding of the issues and assumptions made when comparing means or medians you are advised to explore the following texts.

Practical Statistics for Medical Research, DG Altman, Chapman and Hall 1991. Chapter 9

An Introduction to Medical Statistics, 3rd Ed, M Bland, Oxford Medical Publications 2000. Chapters 10 and 12

Medical Statistics: A textbook for the Health Sciences. 4th Ed, MJ Campbell, D Machin & SJ Walters, Wiley 2007. Chapter 8

Essential Medical Statistics, 2nd Ed BR Kirkwood & JAC Sterne, Blackwell Science 2003. Chapters 7,9 & 30.

Medical Statistics at a Glance 3rd Ed, A Petrie & C Sabin, Wiley-Blackwell, 2009.


28 Analysing Two or more Groups - Split Files

If you want to analyse your data by sub groups within the study sample, Split Files is the command to use. Suppose you want to analyse your data by gender groups, male and female separately.

In the Data Editor window,

Either  Click on the Split file button
or Select **Data | Split File** from the menu.

Click on *Compare Groups*

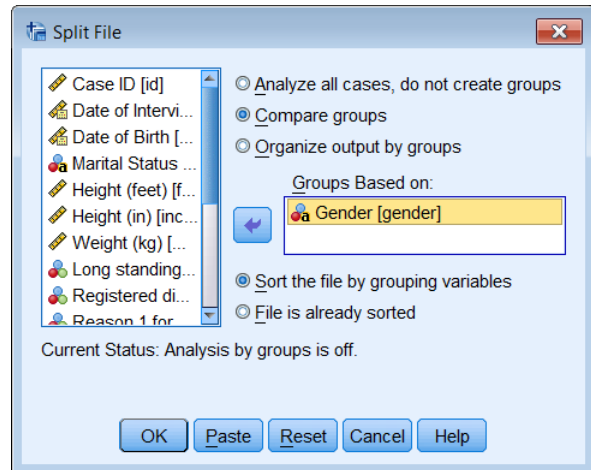
Highlight *gender* in the variable list drag it or press  to add this to the *Groups based on:* panel.

Check that the *Sort the file by grouping variables* option is selected.

Click on **Paste**.

In the syntax window, highlight the commands

SORT CASES BY gender .
SPLIT FILE LAYERED BY gender .



Click on the **Run Selection** button .

All further analyses will be carried out for females and males separately, rather than for all cases.

The difference between the output produced by the options *Compare Groups* and *Organise output by groups* is illustrated below:

Compare Groups

	Frequency
Females	
Males	

Organise output by groups

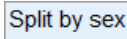
Gender =
Female

Frequency

Gender = Male

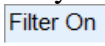
Frequency

Organise output by groups creates more lines of output although the same statistics are produced.

Note: the SPSS status bar changes to show that Split Files is active,  is displayed.

To turn the split files effect off, select **Data | Split Files** or click on  and **Click** on *Analyze all cases, do not create groups*. Click on **OK**, or **Paste** the command SPLIT FILE OFF into the syntax window and run it.

29 Case selection – Filtering

SPSS provides a 'filtering' option for analysing a subset of your cases. When filtering is turned on, as shown on the status bar , only those cases meeting the selection criterion are used for analysis. The criteria used to define sub groups can include:

- Variable values and ranges
- Date and time ranges
- Case (row) numbers
- Arithmetic expressions
- Logical expressions
- Functions

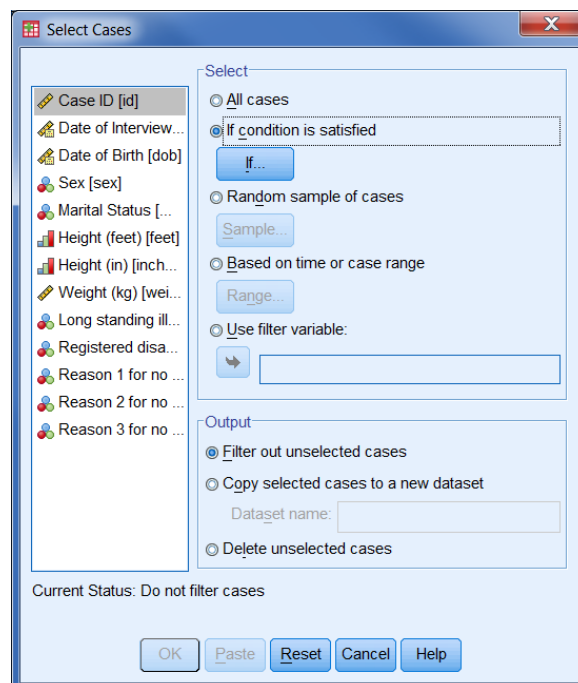
Suppose that you wished to analyse the data for those respondents that have a long standing illness (the variable LSILL will have a value of 1). In the Data Editor window, either:

Click on the Select Cases button 

or

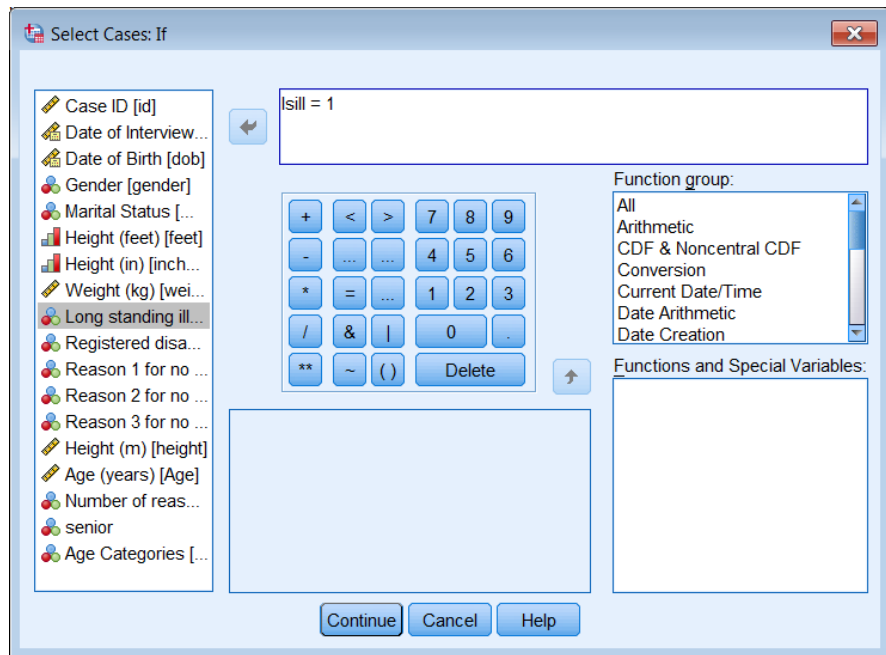
Select **Data | Select Cases** from the menu.

There are four ways of selecting cases, the *If condition is satisfied* option is the most frequently used.



Choose this option by clicking the **If...** button. A *Select Cases: If* dialogue box is displayed.

If we wish to analyse those cases where the variable value is 1, an expression **Isill = 1** is built in the text panel by clicking on the buttons.



Click on **Continue**.

Paste and **Run** the commands in the Syntax window

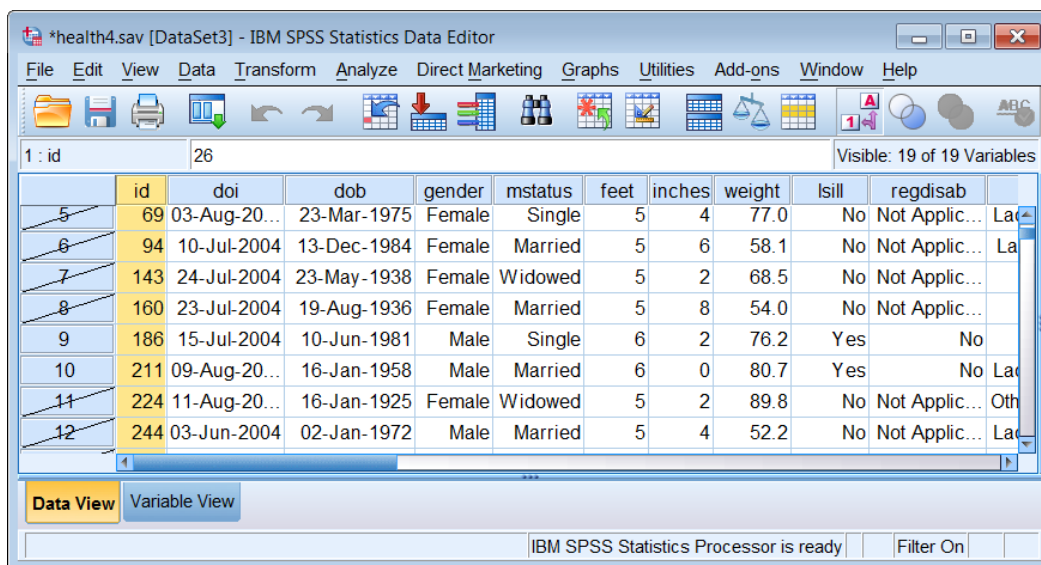
```


USE ALL.
COMPUTE filter_$(=Isill = 1).
VARIABLE LABEL filter_$ 'Isill = 1 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMAT filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE .

```

Filtered (out) cases are left in the data editor with a diagonal line through the case number. A variable, **filter_\$**, is also created, this has a value of 1 for cases that are Selected (10 and 11) and 0 for cases that are Not Selected (7, 8, 9, 12 and 13).

The Status bar in the Data Editor displays **Filter On**.



To turn the Filtering off, click  or select **Data | Select Cases**.
Click on *All Cases*.

Paste and **Run** the commands.

FILTER OFF.
USE ALL.
EXECUTE .

If you want to perform a single analysis on a subgroup, use filtering. But if you want to do intensive work with selected cases, *Delete unselected cases* is probably more efficient. In this case it is worthwhile saving these cases to another SPSS data file.

30 Sorting Cases

The data set can be sorted by cases (rows) using **Data | Sort Cases**.

You can select multiple sort variables, cases are sorted by each variable within categories of the preceding variable on the *Sort by*: list. For example, if you select Gender as the first sorting variable and Weight as the second sorting variable, cases will be sorted by Weight within each Gender category.

The *Sort Order* can be Ascending or Descending for each variable.

